Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Walter V. Bingham, *Washington, D. C.*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Table of Contents

Permanence of Interest Scores Over 22 Years: E. K. Strong, Jr. 89
Hearing Loss and Worker Morale: F. R. Zintz and W. A. Kerr 92
Output Rates among Chocolate Dippers: H. F. Rothe 94
The Prediction of Driving Skill: P. F. Bartelme, E. D. Fletcher, C. W. Brown, and E. E. Ghiselli 98
Variability of Reaction Time and Susceptibility to Automobile Accidents: W. L. Cation, G. E. Mount, and R. Brenner 101
Psychological Climate and Accidents in an Automotive Plant: V. Keenan, W. Kerr, and W. Sherman 108
Industrial Use of the Partington Pathways Test: L. W. Cozan 112
Mirror-Tracing Practice for Medical and Dental Students: N. V. Scheidemann 114
A Study of Interview Refusals: S. Benson, W. P. Booman, and K. E. Clark 116
Social Characteristics of Respondents and Non-Respondents in a Questionnaire Study of Later Maturity: J. Pan 120
MMPI and Student Teaching: J. U. Michaelis and F. T. Tyler 122
Correction for "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the MMPI": O. M. Lough 125
Ortho-Rater Color Vision Test Compared with the Ishihara and the Pseudo-Isochromatic Plates: N. C. Kephart and M. J. Tieszen 127
A Study of Tachistoscopic Vision: N. C. Kephart and J. D. Pecsok 130
An Experimental Consumer Panel Technique: E. A. Fleishman 133
Book Reviews 136
New Books, Monographs, and Pamphlets 140

American Psychological Association

Vol. 35, No. 2 April, 1951

Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

Subscriptions and business communications should be sent to
American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.

Articles for publication and books for review should be sent to the Editor
Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described by J. E. Anderson and W. L. Valentine in "The preparation of articles for publication in the journals of the American Psychological Association," *Psychol. Bull.*, 1944, 41, 345–376. A reprint of this article will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947
Copyright, 1951, by The American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 35, NO. 2 APRIL, 1951

Permanence of Interest Scores Over 22 Years

Edward K. Strong, Jr.

Stanford University

A study of permanence of scores from the Vocational Interest Blank for intervals of time up to ten years was published in 1943 (**2**, Ch. 15). The data were based on test-retests of college seniors, freshmen, and 11th-grade high school students.

To the above are now added test-retest correlations for as long a period as 19 years for college freshmen and 22 years for college seniors.

In addition to answering the question: how permanent are interest scores, the data throw light upon additional questions. First, is interest similar to memory in that as time goes on the agreement between test and retest becomes less? Second, are interests progressively better established as one grows older so that the older one is when tested the better the agreement between that test and a subsequent one? Third, which affects permanency more in the period from 17 to 44 years of age: (a) interval of time or (b) age at time of taking the test?

The data are based on: (1) seniors at Stanford University first tested in 1927 and retested in 1932, 1937 and 1949, and (2) freshmen first tested in 1930 and retested in 1931, 1939 and 1949. Six different intervals of time can be studied as to their effect upon permanency of interest scores from the data respecting seniors and, similarly, respecting freshmen. Aside from the interval of time factor in the test-retests there is also the age factor; the seniors varied from 22 to 32 years of age and the freshmen from 19 to 28 years of age at the time of taking the test. The combinations of time interval and age are set forth in Table 1.

Method of Analysis

There are several different ways of measuring permanence of interest scores but the method used here is that of correlation between the scores constituting the interest profile of the test with the scores of the interest profile of the retest. In other words the 34 occupational standard scores of a man in the test are correlated with the 34 corresponding scores of that man in the retest. Summaries of fifty or more such correlations are given in Table 1. Such correlations indicate how likely it is that the same general deduction from all scores will be valid in the future.

Correlations for men ranging from 17 to 32 years of age and for intervals of time from 3 weeks to 22 years are given in Table 1. The data for the high school students are from Finch (**1**). He used the mean to summarize the correlations from many cases. Our summary is in terms of the median of the coefficients. (With our data medians and means are practically identical, differing on the average by only .005.) In most groups the median is based on the first fifty cases in our file. The samples for seniors are not identical since some men did not take the test in 1932 or 1937; the same is true for the samples of freshmen.

In columns 7 and 8 of Table 1 correlations are given in which the scores of 100 cases have all been plotted in one scatter diagram and one correlation calculated for all the cases. These coefficients average .035 lower than the medians of individual coefficients. It is probable that the data for the few individuals with very low correlations have affected the second set of coefficients to a greater extent than the first set. In any case we believe the median correlation of many individual records is more enlightening than the single correlation based on the intermingling of one hundred records.

Table 1

Permanence of Interest Scores

No. of Subjects (1)	Educational Level (2)	Dates Tested (3)	Age at Time of Test (4)	Interval in Years between Test and Retest (5)	Median Correlation (6)	Number of Subjects (7)	Correlation (8)
33	11th grade	—	17	.06	.86		
148	11th grade	—	17	1.25	.83		
57	11th grade	—	17	2.33	.81		
50	College freshmen	1930–31	19	1	.88	100	.83
50	College freshmen	1930–39	19	9	.67	100	.67
50	College freshmen	1930–49	19	19	.72	100	.66
50	College freshmen	1931–39	20	8	.72	100	.72
50	College freshmen	1931–49	20	18	.72	100	.70
50	College seniors	1927–32	22	5	.84	100	.78
50	College seniors	1927–37	22	10	.82	100	.74
228	College seniors	1927–49	22	22	.75	100	.72
50	College seniors	1932–37	27	5	.86		
50	College seniors	1932–49	27	17	.84		
50	College freshmen	1939–49	28	10	.87	100	.84
50	College seniors	1937–49	32	12	.88		

Results

The correlation coefficients are amazingly high for test-retest cases: .84 for 5 years; .82 for 10 years; and .75 for 22 years with college seniors. A fourth of all the correlations in the six sub-groups of seniors are .90 or higher, up to .994, the highest, and 60 per cent are .80 or higher; see Table 2.

We concluded in 1943 that "approximately the same rank-order is maintained for occupational interest scores for intervals ranging between one and ten years." Now we can add "between one and twenty-two years." In other words, those who had interests most similar to engineer, lawyer, or minister on the first occasion were the ones who had scores most similar to those same criterion groups on the second occasion, and vice-versa. The data establish the fact that there is greater permanence in interest scores among students than has generally been believed.

Table 2

Percentage Distribution of Correlations Shown in Table 1

r's	Six Groups of Seniors	Six Groups of Freshmen
.90 and over	25.7	19.0
.80 to .89	34.4	27.7
.60 to .69	26.3	29.0
0 to .59	13.3	22.0
−.30 to 0	0.3	2.3

We concluded in 1943 that "permanence of interest scores is somewhat less than for intelligence test scores but more permanent over the college period than college grades, and distinctly higher than for attitude test scores." We now wonder whether or not intelligence test scores will prove to be any more permanent than interest scores for periods of ten to twenty years.

The average reliability of the interest scales upon which the scores are based is about .89. On this basis the correlation between test and retest scores on a single scale should not rise above .89. The correlations considered here are not, however, correlations of scores on a single scale, but correlations between scores on 34 different scales on the test and on the subsequent retest. What the reliability is of 34 scores in a profile has not been calculated. It seemingly must be higher than .89 or we would

not have five correlations in Table 1 between .86 and .88 based on cases ranging from 17-year-old boys with an interval of three months to 32-year-old men with an interval of twelve years.

Permanence of interest scores is affected by the two factors of age at the time of taking the test and interval of time between test and retest. From the data in Table 1 it appears that the two factors are of equal importance, for the correlation between permanence and age with time interval partialed out is .818 and between permanence and time interval with age partialed out is .812.

To the extent that permanence is affected by interval of time we may characterize interests, first, as similar to memories which fade with time, second, as phenomena which progressively change with increasing maturity, and third, as phenomena which shift with experiences peculiar to the individual. We need to know how much all changes in interests are caused by maturation, how much they are caused by individual experiences and how much by forgetting.

According to our calculations, permanence of interests is affected as much by the age at which the test is taken as by interval of time. It has previously been shown that maturation explains many changes in interests—that interests change rapidly from those held at 15 years to those held at about 25 years, and then shift in the reverse direction much more slowly from about 25 years to 55 years of age. Of the change from 15 to 25 years, roughly speaking, one-third of the change in interests is between 15.5 and 16.5 years, one-third between 16.5 and 18.5 years, and one-third between 18.5 and 25 years (**2**, Chs. 12, 13).

As the changes in interests because of maturation are relatively small from 22 to 32 years of age we should expect high permanence of interests from men within this age range. This is what is shown in the table, for all the correlations are between .819 and .880 with one exception and that is for 22-year-old seniors retested twenty-two years later. On the other hand, the lowest correlations in the table, i.e., .667 to .725, are for younger men, 19 to 20 years old, retested 8 to 19 years later. It is unfortunate that there are no data of still younger men tested over long periods of time, and of older men tested for very short intervals of time.

At least to the extent that permanence increases with age, interests are not to be associated with memory for memories do not improve as one grows older. If anything, the reverse is the situation. We feel, at least, that it is harder and harder to remember faces and names as we become older. How far early memories are better established than older ones because of youth and how much because of repetition is unknown to the writer.

Evidently, interests become better and better established with increasing age and consequently the older the man when first tested the more likelihood that a test and retest will agree. But the data in the table indicate that although permanence is less for young men than older men, it is still remarkably high for young men. The correlation of .81 for 17-year-old boys over an interval of 2.3 years and of .72 for 19-year-old college freshmen over an interval of 19 years are evidence of high permanency of interests among young men.

Interests are not separate entities but are an aspect of activities, which are habits. As we grow older certain activities are discarded but many more are engaged in continuously. A human being is more than a mere collection of activities, but the sum total of his activities constitutes a vital part of his makeup. It is this part of a man we measure with an interest test. Indirectly an interest test must measure to some extent man's underlying wants and desires, on the one hand, and his abilities, on the other hand, for the activities must represent what he wants to do modified by what he is capable of doing.

In any attempt to test a man in order to discover what kind of person he is it is evident that an interest test should be included in the battery.

Received June 17, 1950.

References

1. Finch, F. H. The permanence of vocational interest. *Psychol. Bull.*, 1935, **32**, 682.
2. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford, Calif.: Stanford University Press, 1943.

Hearing Loss and Worker Morale *

F. Raymond Zintz

Otarion, Inc., Chicago

and

Willard A. Kerr

Illinois Institute of Technology

The purpose of the present study was to determine the magnitude of relationship which exists between job morale and American Medical Association binaural hearing loss. Subjects for the research were 91 per cent of the entire personnel of the (non-union) Chicago plant of the Automatic Spring Coiling Company, a total of 53 personnel cooperating. While the sample is not large, the fact that the sample is almost a total one containing a random collection of levels of adult hearing efficiency is reassuring for this research.

The personnel manager sent each worker a letter announcing the Illinois Institute of Technology research program, stating that all cooperation voluntarily given would be without reference to any worker's identity, and that the audiogram would be presented to each cooperating worker.

Hearing Loss. Binaural hearing loss was computed on each employee according to the specifications of the American Medical Association, after testing in a quiet private office by an Otarion-trained audiometrician with a Model 600 Otarion Audiometer, a standard instrument measuring acuity at each of the conventional frequencies. Only one of the 53 workers wore a hearing aid; the remainder except one apparently regarded themselves as normal in hearing ability. In the entire group, hearing loss ranged from less than one per cent to 75 per cent.

Job Morale. Employing the anonymous (but secretly sequence research-coded) *Tear Ballot for Industry* (1), ten job satisfaction variables were measured on a five-point scale and summated: job security feeling, belief in employee welfare interest of company, attitude toward immediate supervisor, attitude toward working conditions, attitude toward co-workers, attitude toward pay, attitude toward communication freedom, confidence in good sense of management, and opinion of effect of company experience on personal happiness. These scores with a total possible range from 10 to 50 actually ranged from 27 to 48. Age also is reported (in five-year multiples) on the *Tear Ballot.*

In addition to this general job satisfaction score, two sub-scores were computed on each employee—interpersonal satisfaction score and "impersonal" satisfaction score. The interpersonal score consisted of a summation of responses to "How does your immediate superior compare with other managers, foremen, or section leaders as to supervisory ability," "Are most of the workers around you the kind

Table 1

Tetrachoric Intercorrelations Among Certain Hearing Loss and Morale Values of 53 Employees in a Coil Spring Factory*

Variables	2	3	4	5	6
1. Per cent hearing loss (AMA)	.45	−.25	.43	.23	−.16
2. Age		.34	.85	.04	.43
3. Job satisfaction			.31	.68	.86
4. Past mean job tenure				.14	.19
5. Interpersonal job satisfaction					.50
6. "Impersonal" job satisfaction					

Partial correlations:

$r_{13.2} = -.42$ $r_{14.2} = .12$

$r_{15.2} = .24$ $r_{34.2} = .12$

$r_{16.2} = -.46$ $r_{13.24} = -.44$

* Coefficients statistically significant at the five per cent level or better are printed in boldface type.

* The authors acknowledge with gratitude the cooperation in this research of the Automatic Spring Coiling Company, Roy O. Carlson, and James Fairchild.

who will remember you when you pass them on the street," and "Do you feel that you have proper opportunity to present a problem, complaint, or suggestion to the management," as such replies theoretically represent an employee's interpreted success, adaptation, and adjustment in interpersonal relations. The "impersonal" score consisted of a summation of the other seven items of the original ten.

Turnover Rate. A separate information slip requested number of previous civilian employers and also number of years work experience under these same employers. Thus "years" divided by "employers" allowed computation of a past mean tenure (reverse of turnover) value on each worker, as an independent estimate of the worker's past adjustment. These mean past tenures, admittedly responsive to myriad factors other than adjustment, ranged from half a year to 24 years.

Intercorrelations and some partial correlations were computed to exploit the obtained data, and these are presented in Table 1.

Results

As indicated in Table 1, the various criteria of worker morale are contaminated with the age factor. When the age factor is held constant by partial correlation, a statistically significant negative relationship survives between hearing loss and total job satisfaction. The magnitude of this correlation (−.42) suggests that about one-sixth of the variance in job satisfaction is accounted for by variance in hearing ability. When the total *Tear Ballot* score is broken into the interpersonal relations and the relatively "impersonal" sections, the interpersonal job satisfaction section fails to correlate significantly with hearing loss when age is held constant; this is an implied contradiction of the clinical theory (2) that hearing loss is specifically a paranoid pitfall. The remaining more "impersonal" job satisfaction value correlates −.46 with hearing loss when age is held constant.

Mean tenure in past jobs is unrelated to hearing loss when age is held constant. When both age and mean past tenure are held constant, total *Tear Ballot* job satisfaction still bears a relation of −.44 with AMA hearing loss.

Summary

Fifty-three (91 per cent of total personnel) employees of a small factory were measured for AMA binaural hearing loss, job satisfaction, mean tenure in past jobs, and age, in an attempt to measure some of the possible psychological effects of hearing loss on worker morale.

1. When age is held constant by partial correlation, variance in hearing loss appears to account for one-sixth of the variance in job satisfaction ($r = -.42$). Hearing loss may be considered an important factor in job morale.
2. Attitudes towards associates seem to be much less influenced by hearing loss ($r = .24$) than are the more "impersonal" aspects of job satisfaction ($r = -.46$).
3. Mean tenure of past jobs is unrelated to hearing loss when age is held constant.

Received May 5, 1950.

References

1. Kerr, W. A. On the validity and reliability of the job satisfaction tear ballot. *J. appl. Psychol.*, 1948, **32**, 275–281.
2. O'Kelly, L. I. *Introduction to psychopathology.* New York: Prentice-Hall, Inc., 1949, p. 261.

Output Rates among Chocolate Dippers

Harold F. Rothe

Stevenson, Jordan & Harrison, Inc., Chicago, Illinois

There is a growing recognition among psychologists of the difficulties of obtaining adequate criteria in industrial work. An increasing amount of space is being devoted to this topic in recent books and conferences. These difficulties are as true of objective performance data as of subjective criteria such as merit ratings. In a recent paper, Einar Hardin made particularly clear the difficulties involved in obtaining an adequate performance criterion.[1]

Background of the Problem

In a previous paper the writer presented data on the output of butter wrappers, showing that a large number of daily work curves (based on production during each fifteen-minute period) must be obtained before a stable or consistent curve for an operator could be obtained. Daily work curves for individuals had but little consistency from day to day (**1**). These butter wrappers showed an enormous range of productivity from one fifteen-minute period to the next, so great that the variations within one operator's performances over a five-day period, were greater than the average variation from one operator to another over the same period (**2**).

In another paper on machine operators it was shown that the operators had relatively little consistency from one two-week period to the next, these correlations being .57 and .68 in that study (**3**).

On the other hand, Tiffin has described a week to week consistency of .96 for hosiery loopers (**4**). It is probably noteworthy that the hosiery loopers were paid on an incentive system, and the butter wrappers and machine operators were not. It is also probable that other factors such as shortages of material, good or poor scheduling, amount of work ahead of each operator, validity of recorded data, and even more subtle factors affect the consistency of the operators from period to period. It is certainly clear that a great deal more detailed data must be obtained, together with a complete description of the setting in which the study took place, before industrial psychologists can reach sound conclusions regarding output data.

The purpose of the present paper is to present another study, and a description of the setting of that study, as another step in the development of an understanding of output rates.

Background of the Study

The data for this study were taken from the official books of the Fannie May Candy Company of Chicago, Illinois.[2] They cover the period December 16, 1949 through April 15, 1950. They refer to regular, experienced female employees, working at their regular jobs at their usual workplaces. This job is described in the USES Dictionary of Occupational Titles, code number 6-05.312.

The Fannie May Candy Company is characterized by friendly employee-management relations. There has long been in effect an annual bonus system, a pension plan, a hospital plan, and similar aspects of progressive employee relations. There is an employee representative committee that represents the employees. Further, each employee feels free to walk into the General Manager's office at any time.

There is an incentive system that has been in effect for many years, and the employees are paid a one-to-one ratio for performance over standard. It should be noted particularly in the data that follow here that this incentive system is an *effective* one, that is, the employees actually *do* regularly produce more than 100 per cent of performance, and get paid accordingly. Finally, this job is completely controlled by the operator; there are no moving

[1] A paper presented at the Sixth Annual Conference on Industrial Relations Research held at the University of Minnesota, May 9 and 10, 1950.

[2] The writer wishes to thank Mr. Harry H. Simpson, Vice-President of the Fannie May Candy Company, for permission to publish these data.

belts or machinery that govern the pace of the hand dippers.

Since this study was limited to experienced hand dippers, and does not include persons on other jobs, part-time dippers, or relatively inexperienced dippers or night-shift dippers, the sample is small, being only 18 girls. No data were included for any girl who did not work at least one full day on dipping in any one week, and, hence, the size of the sample dropped occasionally to a smaller number, but on no occasion were there less than 15 girls included in the data for any one week.

Table 1

Weekly Average Output (Per Cent Performance of Standard) for Group of Chocolate Dippers

Week Ending	Per Cent Performance	Number of Girls Weekly
December 31	128.4	16
January 7	127.5	18
14	129.5	18
21	128.8	17
28	129.0	17
February 4	126.9	18
11	131.5	17
18	127.5	15
25	127.9	16
March 4	126.7	15
11	127.2	17
18	127.2	17
25	132.7	16
April 1	134.2	16
8	128.4	16
15	126.0	16

Data

Analysis of the trend of average weekly performance of the chocolate dippers shows that there was no consistent long-term upward or downward trend in productivity during this period. These data are shown in Table 1. Thus, it may be concluded that the period of time studied here was an essentially "normal" one.

The correlation of each operator's performance for one week with her performance for the following week was determined by the rank difference method. The distribution of the obtained Rho's is shown in Table 2. The median inter-weekly correlation is .85. Thus each girl was quite consistent in her performance from week to week.

TABLE 2

Frequency Distribution of Rho's between Successive Weeks' Output Individual Performance for Group of Chocolate Dippers

Rho	Frequency
.96–1.00	2
.91– .95	2
.86– .90	3
.81– .85	2
.76– .80	2
.71– .75	2
.66– .70	0
.61– .65	1
.56– .60	1
Median Rho = .85	

The greatest and least amount of productivity for each girl for any one of the 16 weekly periods is shown in Table 3, together with the ratio of best to worst performance for each girl.

Table 3

Highest and Lowest Average Weekly Performances, and their Ratios, for Individual Chocolate Dippers During 16-Week Period

Girl	Highest Weekly Average	Lowest Weekly Average	Ratio of Highest to Lowest
A	135	116	1.16
B	123	95	1.29
C	142	120	1.18
D	175	144	1.22
E	134	108	1.24
F	114	104	1.10
G	131	103	1.27
H	149	128	1.16
I	146	132	1.11
J	136	125	1.09
K	138	108	1.28
L	150	136	1.10
M	124	114	1.09
N	145	117	1.24
O	140	126	1.11
P	147	103	1.43
Q	140	113	1.24
R	133	120	1.11
S	152	127	1.20
Median intra-individual ratio = 1.18			

Table 4

Highest and Lowest Average Individual Weekly Performances, and their Ratios, for Group of Chocolate Dippers During 16-Week Period

Week Ending		Highest Girl's Average	Lowest Girl's Average	Ratio of Highest to Lowest
December	31	147	105	1.40
January	7	149	101	1.48
	14	150	102	1.47
	21	152	107	1.42
	28	150	99	1.52
February	4	149	99	1.51
	11	175	101	1.73
	18	144	98	1.47
	25	157	95	1.65
March	4	155	98	1.58
	11	156	110	1.42
	18	156	106	1.47
	25	150	113	1.33
April	1	159	104	1.53
	8	157	101	1.55
	15	148	103	1.44

Median inter-individual ratio = 1.475

The median performance ratio of best to worst for these girls is 1.2. Thus, girl A produced a high of 135 per cent for one of these 16 weeks, and her lowest weekly average production during this time was 116 per cent, giving her a best/worst, or intra-individual ratio of 1.16.

The ratio of best operator to worst operator, for each week, is shown in Table 4. During the week ending December 31, 1949, one girl produced 147 per cent and another girl produced 105 per cent, for an inter-individual ratio of 1.40.

From Tables 3 and 4 it can be seen that the average ratio of the range of inter-individual performance is greater than the average ratio of the range of intra-individual performance.

Discussion

The present writer has previously stated two tentative hypotheses relating to the effectiveness of incentives. One was that "the incentives to work may be considered ineffective when the ratio of the range of intra-individual differences is greater than the ratio of the range of inter-individual differences" (**2**, p. 326). It is quite apparent that here, where we may safely conclude that the various incentives were effective, that this hypothesis is supported by these data.

The second hypothesis was that "if the intercorrelation of group output rates for two periods closely related in time is less than .80 the incentivation is not highly effective, while intercorrelation higher than .90 indicates effective incentivation" (**3**, p. 488). This hypothesis is not supported by this study, although the error is in amount, not in direction. It is, of course, risky business making a hypothesis with as little data as exist in this field. On the other hand, hypotheses may stimulate someone else to do research (and to disprove them).

On the strength of other data which have not been published, the writer believes that there is a relationship between consistency of data and effectiveness of incentives. He also believes that the incentivation of the machine operators (previously described) was actually fairly high (**3**), even though there was no incentive system there at the time. That is, high incentivation probably existed there because of an intangible attitude and in spite of the lack of an engineered incentive system as such at the time of that study.

Hence, the second hypothesis is herewith changed to read, "if the intercorrelation of output rates for two periods closely related in time is less than .50 the incentivation is not highly effective, while intercorrelation higher than .80 indicates very effective incentivation."

This hypothesis is of course tentative. It is also likely that these coefficients will vary with the length of the periods of time correlated. Presumably lower correlations will be found if short periods of time are correlated than if long periods are correlated.

Another point is also important here. It is apparent that the writer is attempting to demonstrate something other than "reliability" by proposing these hypotheses. That is, it appears that the consistency of output, from time to time, may reflect a real phenomenon and permit some prediction (such as an estimate of the effectiveness of the incentives in a situation) and not merely reflect the "reliability" or "unreliability" of the criterion. The concept of "reliability" has been taken over from the

realm of testing where it is related primarily to errors of measurement. Such errors undoubtedly exist in work such as is described in this article. But it is also very likely that there is a "consistency" or "inconsistency" of industrial output data that is a real phenomenon and not merely an error of measurement.[3]

Summary

Analysis of the output rates of 18 chocolate dippers for a sixteen-week period, under conditions of excellent incentivation, reveals an average week to week consistency coefficient of .85 and an average ratio of the range of inter-individual performance greater than the average ratio of the range of intra-individual performance. Both these findings tend to support some hypotheses previously stated by the writer, although with changes in one of the hypotheses.

The probability that coefficients of consistency of output rates, from time to time, reflect some real phenomenon or co-variate, and are not merely indicative of errors of measurement ("reliability") is discussed.

Received June 10, 1950.

[3] This point has also been expounded by Einar Hardin, previously cited. It is sincerely hoped that he will soon publish his paper on the subject.

References

1. Rothe, H. F. Output rates among butter wrappers: I. *J. appl. Psychol.*, 1946, **30**, 199–211.
2. Rothe, H. F. Output rates among butter wrappers: II. *J. appl. Psychol.*, 1946, **30**, 320–327.
3. Rothe, H. F. Output rates among machine operators: I. *J. appl. Psychol.*, 1947, **31**, 484–489.
4. Tiffin, J. *Industrial psychology.* New York: Prentice-Hall, 1942.

The Prediction of Driving Skill

Phyllis F. Bartelme
Institute of Welfare Research, New York City
Edwin D. Fletcher
Department of Motor Vehicles, California
and
Clarence W. Brown and Edwin E. Ghiselli
University of California

The most frequently utilized index for gauging proficiency in the operation of motor vehicles is accident rate. As a consequence, in most instances the validity of tests for use in the selection of motor vehicle operators is determined by relating scores to some measure of safety. Since the supposition is that to a considerable degree accidents are due to lack of driving skill, selection tests largely are of the apparatus type designed to measure basic motor and perceptual capacities.

Two of the present authors recently considered the problem of the usefulness of paper and pencil tests in the prediction of accidents of motor vehicle operators (1). Scores on a paper and pencil test battery, consisting of tests of dotting, tapping, linear distance discrimination, judgment of distance through perspective and interposition, and mechanical principles, were found to be significantly related to accidents among taxi cab drivers. This battery previously was found to be effective in the prediction of the number of accidents incurred by street car motormen. Since this battery had proved to be effective it was argued that at least for selective purposes paper and pencil tests might be devised which would be as effective as the more expensive and cumbersome performance tests.

One shortcoming of paper and pencil tests lies in the fact that while the scores they yield might be predictive of the degree of safety in performance, they are of far less value than performance tests for purposes of diagnosing particular deficiencies which might well lead to accidents. Knowledge of such deficiencies would be most helpful in remedial training. The possibility remains, however, that paper and pencil tests might be useful in indicating which individuals are deficient in driving skill. By retraining in basic driving skills operating safety presumably would be improved. It was the purpose of the present investigation to determine the validity of the aforementioned paper and pencil tests in the prediction of driving skill. If it appears that the validity is significant then such tests might be utilized as part of a diagnostic battery.

Methods and Procedure

An opportunity arose in a large-scale driver testing program for comparing the paper and pencil tests with two of the most frequently used driver performance tests in terms of their relationships with work-sample tests involving the actual operation of motor vehicles. Two of the authors (Bartelme and Fletcher) conducted a drivers clinic at one of the larger army bases in which the driving skills of over two hundred operators were evaluated. In addition to being examined on the paper and pencil tests the operators were examined on the steering and brake reaction time test in use by the California Department of Motor Vehicles in their drivers clinic.

After completing these tests the drivers were given a modified version of the exercises for driving skill developed under the sponsorship of the American Automobile Association (2). There were four of these work-sample exercises as follows:

1. Driving forward and backward on a straight line for a distance of 100 feet.

2. Driving forward and backward in a narrow lane affording only about two feet clearance on each side.

3. Aligning the front and back wheels and the front and back extremities of the car with

Table 1

Coefficients of Correlations between Various Tests and Total Score on the Work-Sample Driving Exercises

Type of Test	Operators of Light Vehicles		Operators of Medium Vehicles		Operators of Heavy Vehicles		All Operators	
	$\bar{r}$*	N	$\bar{r}$	N	$\bar{r}$*	N	$\bar{r}$*	N
Dotting	.03	66	.08	58	−.16	70	−.02	194
Tapping	.10	65	.13	58	−.05	70	.06	193
Distance Discrimination	.46	44	.01	36	.16	74	.21	154
Mechanical Principles	.11	43	.24	38	−.03	73	.08	154
Judgment of Distance	.08	44	−.01	37	−.05	71	−.01	152
Paper and Pencil Test Battery	.24	37	.11	34	−.12	56	.05	127
Steering	.23	72	.20	58	.18	87	.20	217
Reaction Time	−.11	42	.11	34	.09	52	.02	128

* These coefficients are averages from two or more vehicle groups, using the Fisher $\bar{z}$ method.

crossmarks on the roadway, both forward and backward movement of the car being used.

4. Brake reaction time (movement of foot from accelerator to brake), and stopping distance using the detonator test.

Each driver was allowed to perform the foregoing exercises with the type of car he customarily drove in his employment. The performance on each exercise was scored in terms of judgments of both qualitative and quantitative characteristics. The total score consisted of a simple sum of the point scores of the four exercises. This score was used in evaluating the several tests.

Results

Table 1 presents the coefficients of correlation between the total scores on the work-sample driving exercises and scores on the various tests. Under light vehicles are included passenger cars and ¾ ton trucks. Medium vehicles were 1½ ton trucks. Heavy vehicles included 2½ ton trucks and buses. When two or more types of vehicles were combined, the resulting coefficient was obtained by averaging the coefficients for the several types through use of the Fisher $\bar{z}$ method.

It will be noted that the individual paper and pencil tests show little promise of predicting actual performance. The Distance Discrimination Test gave the highest coefficients. This result can be rationalized in terms of the visual characteristics involved in the particular exercises used in the work-sample tests. The Distance Discrimination Test would seem to merit further study in driver testing situations.

Although the evidence can be considered merely suggestive, it would appear that the relationships between the paper and pencil tests and the work-sample tests is not the same for heavy vehicles as for light vehicles. In the case of heavy vehicles, the greater number of negative coefficients, even though they are of small size, suggests the possibility of different types of functions underlying performance on light and heavy motor vehicles.

Of the two performance tests the steering reaction test is superior to the brake reaction time test. The steering test gave consistently positive results for the various kinds of vehicles.

The paper and pencil tests were combined in a battery, the weights being those determined in previous investigations with vehicle operators (1). For light cars the paper and pencil test battery shows as high a relationship with the work-sample performance as does the steering test, but is inferior to this test in terms of relationship with performance on the heavier vehicles. It would appear that insofar as the tests used in the present study are concerned, the advantage lies with the performance tests, and particularly with the steering test.

Conclusions

It appears from the findings of the present investigation that the prediction of driving skill by means of tests is by no means wholly satisfactory. This conclusion is consistent with the

results of other investigations (3). It can be concluded that insofar as driving ability was measured at all by the tests used, the steering test consistently gave the best results for all types of automotive drivers, and that for operators of light vehicles a battery of paper and pencil tests was equally good.

Summary

Achievement scores of operators of motor vehicles on five paper and pencil and two performance tests were compared with their scores on work-sample driving tests involving the actual operation of motor vehicles. For the most part the coefficients of correlation were low. A steering performance test gave the most consistent results. With the exception of drivers of light vehicles, paper and pencil tests were found to give inadequate predictions of driving skill.

Received May 29, 1950.

References

1. Ghiselli, E. E., and C. W. Brown. The prediction of accidents of taxicab drivers. *J. appl. Psychol.*, 1949, **33**, 540–546.
2. Neyhart, A. E. *Check-lists and score sheet for exercises for driving skill.* American Automobile Assn., 1944.
3. Personnel Research Section, Adjutant General's Office. *Statistical manual.* 1944.

Variability of Reaction Time and Susceptibility to Automobile Accidents *

W. L. Cation, G. E. Mount, and R. Brenner

Institute of Transportation and Traffic Engineering, University of California, Los Angeles

Reactions to visual and auditory stimuli are made more or less continuously by the automobile driver, particularly on a heavily-traveled street or highway. Individual differences exist in the speed with which drivers are able to respond to these stimuli. Measurements of reaction time have been traditionally included in experimental test batteries designed to predict accident susceptibility of drivers, on the premise that the number of accidents or the accident-rate of an individual driver is partially a function of his reaction time.

However, the reported correlations between reaction time and accident rate have been disappointingly low. One explanation which has been advanced for these low correlations is that the "slow reactor" soon becomes aware of his "handicap" and makes adjustments or compensations accordingly in his driving habits. Another explanation is that it is *variability* of reaction time which is the reaction-time factor related to accident-rate rather than reaction time per se. In other words, it may be that the driver who is *consistent* in the speed with which he reacts to a stimulus, regardless of how high or how low his mean reaction time may be, is least susceptible to accidents.

The purpose of this study was to test the hypothesis that an individual whose reaction times are relatively consistent throughout a short series of measurements, regardless of the magnitude of his mean reaction time, will be found to be less susceptible to automobile accidents than is the individual whose reaction times are more variable.

A review of the published studies since 1930 has shown them to be inconclusive on one or more points.

Johnson (3), in his review of the literature pertaining to the detection and treatment of accident-prone drivers, refers to an unpublished study by Cobb (1) in which a correlation coefficient of .06 ± .03 was obtained between variability of reaction time and propensity to head-against-rear accidents or passing collisions over other types of accidents. In Cobb's study, measurements were made of simple reaction to an auditory stimulus using a sample of 3,663 drivers licensed in Connecticut. Thirty-six measurements were made per subject. The criterion data were secured from the state records of accidents. Johnson states that "the records are not complete, except with respect to fatal accidents. It is probable that not more than half the accidents which are reportable under the law, but which do not involve deaths or personal injury are actually reported, unless the drivers are insured." Johnson also criticizes Cobb's sample on the basis that "it is non-typical of the general driver population of the state, for it contained many drivers whom the authorities had sent in for examination because they had recently got into trouble." As Johnson states, this no doubt tended to bias the distribution and limits the interpretation which may be made of the results obtained. The particular measure of variability which Cobb employed in his study is not given.

In an independent investigation reported by Cobb (1), A. R. Lauer secured a correlation coefficient of −.12 between average variation in reaction time and accident-rate of commercial drivers. It is not stated whether the stimulus employed was visual or aduitory nor whether the reaction was simple, complex or choice. The number of reaction-time measurements recorded for each subject is not given. Lauer examined 1,014 commercial drivers in his investigation. Accident records of the drivers were obtained from the drivers' own statements and from company records. No public records were employed. Accident-rate was established by dividing total number of accidents by the longest driving period given by the driver.

* The writers gratefully acknowledge the cooperation of Capt. N. A. McQuown and Sgt. F. L. Crewe of the Traffic Education Unit, Los Angeles Police Department, in conducting this study.

Lauer attributes the low correlations obtained in his study to the weakness of accident records as a basis for the criteria and to the "multiplicity of causes to which accidents may be ascribed."

Procedure

Simple reaction-time measurements were made by means of a reactometer. The reactometer consists essentially of a chassis on which are mounted a seat, steering wheel, clutch pedal, brake pedal and accelerator pedal. A small, red stimulus light appears before the subject on a screen which is an integral part of the apparatus. The screen is approximately five feet from the subject's eyes as he is seated behind the steering wheel. The accelerator pedal and brake pedal are connected electrically to a timer. The timer records, in hundredths of a second, the time required for the subject to move his foot from the accelerator pedal to the brake pedal in response to the red stimulus light. The stimulus light appears at twelve different instances throughout a total testing period of 80 seconds. The time-intervals between appearances of the stimulus light are not of equal duration. This minimizes the possibility of the subject anticipating the appearance of the stimulus light. The first 20 seconds of the testing period are for practice purposes. The stimulus light appears twice during this practice period, but these two reaction times were not recorded in this study. The remaining ten reaction times were recorded and they represent the basic experimental test data used in the study.

The reactometer test was administered to 104 Radio Patrol Officers of the Patrol Bureau, Los Angeles Police Department, during the months of September, October and November, 1948. This sample of 104 officers was used in computing the statistical reliability of the test data and in computing correlation coefficients between age and mean reaction time, between age and standard deviation of reaction time and between mean reaction time and standard deviation of reaction time.

A second sample was selected from the original sample of 104 officers for the purpose of computing correlation coefficients between the criterion data and the three variables, age, mean reaction time and standard deviation of reaction time. This second sample was selected on the basis of length of service with the Police Department. Since accident records were to be used as the basic criterion data, it was necessary that each officer in this second sample be a member of the Police Department throughout the selected criterion period. The second sample so selected consisted of 87 officers.

Data describing the age and the length of service of each of the two samples are summarized in Table 1. Means and standard deviations shown in Table 1 were computed from ungrouped data.

The indices considered for each subject in this study were the mean of his ten reaction-time measurements and the standard deviation of these measurements. The reliabilities of these indices were computed by the split half technique from alternate raw scores. That of the means was thus found to be .87 which corrected for the number of measures yielded a reliability coefficient of .93. The reliability of the standard deviations was .27 which similarly corrected yielded a reliability coefficient of .41.

The mean and standard deviation of the population of mean reaction times of individual subjects were calculated for both the sample of 104 officers and for the sample of 87 officers. Similarly, the mean and standard deviation of the population of the standard deviations of individual subjects were calculated for each of the two samples. In other words, these latter calculations yielded a "mean of the mean reaction times of the subjects" and a "mean of the standard deviations of reaction times of the subjects," as well as a standard deviation in each of the two instances. These statistics are given in Table 2. All calculations were made from ungrouped data.

Table 1

Statistical Description of Samples

Variable	N	Mean	Standard Deviation	Range
Age	104	34.2	7.8	22–55
	87	35.9	7.4	24–55
Years in Department	104	7.5	5.5	.5–24.0
	87	8.7	5.2	2.0–24.0

Table 2

Mean, Standard Deviation and Range of Mean Reaction Times and Standard Deviations of Reaction Time*

Variable (Based on 10 Reaction-Time Measurements)	N	Mean	Standard Deviation	Range
Mean Reaction Time	104	43.1	4.3	34.8–56.3
	87	43.3	4.5	34.8–56.3
Standard Deviation of Reaction Time	104	3.93	1.53	1.56–9.83
	87	3.87	1.43	1.56–9.61

* Data are shown in hundredths of a second.

The range of the standard deviations shown in Table 2 (1.56–9.83) indicate that differences do exist in the time consistency with which individuals react to a visual stimulus and that those individual differences are of considerable magnitude.

The basic criterion data consisted of the official reports of accidents sustained by the officers while on duty as drivers in police radio patrol cars. Police Department regulations require that each accident in which an officer is involved be reported regardless of how minor it may be. This gave reasonable assurance that the accident reports secured for each officer in the sample represented a complete record of his accidents. Accidents in which a patrol car is involved are investigated and reported upon by an accident investigation unit of the Police Department.

The criterion period extended from January 1, 1947 to January 31, 1949, inclusive. During this period the sample of 87 officers sustained a total of 34 accidents. Sixty-two officers, representing 71 per cent of the sample, had no accidents during the criterion period. In other words, 25 officers, or 29 per cent of the sample, were responsible for the 34 accidents.[1] One of these 25 officers had three accidents, each of seven had two, and the remainder had one each. A test outlined by Mintz and Blum (4) for examining significance of differences between distributions of multiple accidents and the expected Poisson distribution was performed. The results indicated that the obtained distribution did not differ significantly from chance and hence that no conclusion as to "accident proneness" of any of the subjects was justified.

A subjective analysis was made of each accident report and each accident was classified in one of the three different categories defined below:

"A" accident—an accident in which the officer was apparently at fault; an "avoidable" accident.

"B" accident—an accident in which responsibility was doubtful or indeterminable.

"C" accident—an accident in which the officer was apparently not at fault; an "unavoidable" accident.

This analysis resulted in 13 accidents being classified as "A" accidents, 8 being classified as "B" accidents and 13 being classified as "C" accidents. Two of the officers who had more than one accident had one each "A" and "C"; 2 had one each "A" and "B"; 2 had one each "B" and "C"; 1 had two "A's"; and 1 had an "A," a "B" and a "C."

Accident-rate was not computed for each subject since exposure to accidents, in terms of miles and hours driven, time of day on duty and general driving environment, was considered to be sufficiently homogeneous between the four groups of subjects in the sample. The four groups are comprised of the three accident groups and the no-accident group. The groups are mutually exclusive since the individuals having more than one accident were assigned to only one group with an "A" accident taking precedence over a "B" which in turn took precedence over a "C."

Correlations and intercorrelations between the variables were calculated by the tetrachoric technique described by Garrett (2).[2] In constructing the tetrachoric tables the line of division was established at the mean for each of the three variables, age, mean reaction time and standard deviation of reaction time. Three different "splits" were made of the accident

[1] It is recognized that the over-all exposure to which drivers of police patrol cars are subjected is not comparable to that of the general driver population.

[2] This correlation technique was considered appropriate as an initial method of analysis in this particular study. If results obtained by application of the tetrachoric technique are not statistically significant, there is little, if any, justification for applying a more refined technique of analysis.

Table 3

Basic Test Data

Group	Mean Reaction Time	Mean SD Reaction Time
A	40.9	3.54
B	46.76	3.48
C	44.41	3.79
No Accident	43.3	3.97

data. These three "splits" are defined as follows:

1. "No" accidents *vs.* "A," "B" or "C" accidents.
2. "B," "C" or "No" accidents *vs.* "A" accidents.
3. "C" or "No" accidents *vs.* "A" or "B" accidents.

Further analyses, as indicated by the results obtained with the tetrachoric r technique, were made using the analysis of variance technique described by Snedecor (4).

The SE of a tetrachoric coefficient of zero for N = 87 is .17. For N = 104 the SE is .15.

Only one of the nine validity coefficients shown in Table 4 is statistically significant. The correlation of −.30 between mean reaction time and accidents ("split" on the basis "B," "C" or "No" accidents *vs.* "A" accidents) is significant at the five per cent level of confidence. This would indicate that the individuals with short reaction times are more likely to have accidents. However, the only other statistically significant coefficient in Table 4 is the one between mean reaction time and age (r = .44). It follows that the apparent mean reaction time versus accident experience correlation actually could be partly due to an age effect and not entirely to a reaction-time consideration. The correlation of −.18 ± .16 between age and the particular accident split, admittedly not significant statistically, is nevertheless indicative of such a possibility. The computed partial correlation of mean reaction time with accidents is reduced to −.25 with the age factor removed. This −.25 correlation is not significant at the 5 per cent level of confidence. This establishes the need for careful consideration of the results before generalizing as to a mean reaction time–accident experience relationship.

Table 4

Tetrachoric r's and SE's

Variable	Mean Reaction Time*	Age	Accidents: "No" *vs.* "A," "B" or "C"	Accidents: "B," "C" or "No" *vs.* "A"	Accidents: "C" or "No" *vs.* "A" or "B"
SD of Reaction Time*	.22 ± .14 N = 104	.16 ± .14 N = 104	−.07 ± .16 N = 87	.08 ± .16 N = 87	−.04 ± .16 N = 87
Mean Reaction Time*		.44 ± .12 N = 104	.09 ± .16 N = 87	−.30 ± .15 N = 87	−.04 ± .16 N = 87
Age			−.10 ± .16 N = 87	−.18 ± .16 N = 87	−.06 ± .16 N = 87

* Based on 10 reaction-time measurements.

Results

The basic test data are shown in Table 3. The tetrachoric r's and the SE of each r are shown in Table 4.[3]

It is also interesting to note the lack of statistically significant correlation between mean reaction time and standard deviation of reac-

[3] The SE's shown in Table 4 were estimated by applying the formula:

$$SE_{rt} = \frac{1.5(1 - r^2)}{\sqrt{N - 1}}.$$

According to Garrett (2), "an approximation to the SE of a tetrachoric r may be found in the following way: the σ_{rt} is about 50% higher than the SE of an equivalent product-moment r. . ."

Table 5

Analysis of Variance in Mean Reaction Times of Accident Groups "B," "C" and "No"

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Between Groups "B," "C," "No"	58.672	2	29.336
Within Groups "B," "C," "No"	1501.258	72	20.851
Total	1559.930	74	

$$F = \frac{29.336}{20.851} = 1.41$$

$F_{critical}$ 1% = 4.92

5% = 3.13

tion time. This indicates that these two measures are independent variables. This lack of correlation between mean reaction time and standard deviation of reaction time should be taken into consideration in comparing the validity coefficients obtained for mean reaction time with those obtained for standard deviation of reaction time. Apparently, the general similarity between these two sets of validity coefficients is not due to a high correlation between mean reaction time and standard deviation of reaction time. Although mean reaction time apparently increases with age, as indicated by the correlation of .44 between these two variables, variability of reaction time does not, as indicated by the lack of statistically significant correlation between age and standard deviation of reaction time.

Inasmuch as a statistically significant tetrachoric correlation (5 per cent level of confidence) was found between mean reaction time and accidents "split" on the basis "B," "C" or "No" accidents *vs.* "A" accidents, the accident groups "B," "C" and "No" were subjected to a further analysis by applying an analysis of variance technique. Results of this analysis are shown in Table 5.

From Table 5 it can be seen that there is no statistically significant difference between the mean reaction times of accident groups "B," "C" and "No." This established the homogeneity of these three accident groups with respect to mean reaction time and permitted treating them as one group in subsequent analyses.

It was next desired to test the hypothesis that no difference existed between the mean reaction time of accident group "A" and the mean reaction time of accident groups "B," "C" and "No" treated as a single group, with no hypothesis assumed as to population variances. A modified Student's *t*-test, described by Snedecor (**5**, p. 83), was used in which the t' is weighted according to the separately computed variance of each of the two samples. Results of this analysis are shown in Table 6.

From Table 6 it can be seen that there is a statistically significant difference (5 per cent level of confidence) between the mean reaction time of accident group "A" and that of accident groups "B," "C" and "No," when the latter are treated as a single group, with no hypothesis assumed as to population variances. This result is in agreement with the also statis-

Table 6

Significance of Difference between Mean Reaction Time of Accident Group "A" and Mean Reaction Time of Accident Groups "B," "C" and "No" Treated as a Single Group

Accident Group	N	Degrees of Freedom	Group Mean	Sum of Squares	Mean Square	Weighted Mean Square	$t_{.05}$
"B," "C," "No"	75	74	43.7	1559.93	21.08	.28107	1.992
"A"	12	11	40.9	140.88	12.81	1.06750	2.201
	87	85	D = 2.8			1.34857	

$$SD = \sqrt{1.34857} = 1.161 \qquad t = \frac{2.8}{1.161} = 2.41$$

$$t_{weighted\ 5\%} = \frac{(.28107)(1.992) + (1.0675)(2.201)}{1.34857} = 2.16$$

tically significant tetrachoric correlation of −.30 obtained between mean reaction time and the same accident groupings.

Although the tetrachoric technique appears to be useful as a rough, preliminary method for looking at a given body of data from a number of different aspects, two additional points must be recognized before conclusions are drawn. First, this particular tetrachoric correlation was the only significant one of the nine which were calculated, and could have occurred by chance much as the so-called "nonsense correlation" (which is obtained when a body of data is examined from enough different aspects until a significant relationship is found) can exist due to chance. Second, the usual procedures for the evaluation of the statistical significance of a correlation coefficient are not applicable when the correlated data are not normally distributed; this is usually the case with accident records. The commonly used standard errors are then hard to interpret. On the other hand, the reaction times of policemen are probably approximately normally distributed, and thus the *t*-test seems to be entirely appropriate. In this particular case, the obtained agreement between the tetrachoric and the *t*-test simply corroborates the significant relationship indicated by the tetrachoric. While the result can in no way be construed as being conclusive evidence for the existence of this relationship, it is suggestive of such a possibility, and it would be of interest to test a new body of similar data for this relationship.

Conclusions

It can be concluded from the foregoing results that the stated hypothesis has not been established under the conditions of measurement. However, the following lines of investigation should be pursued further before the hypothesis is conclusively rejected:

1. The single kind of measure of reaction time used in this study may not yield variability measurements which are representative of reaction-time situations in general. Basic research is needed to determine the number of different situations for which reaction-time variability must be measured in order to adequately represent general reaction-time variability.

2. It is possible that the subjects used are not representative of other groups of drivers for determining the relationship between reaction-time variability and accidents. Basic research is needed to establish representative classes of drivers within the total population of drivers.

3. The analysis of variance establishes an apparent homogeneity of reaction times of a population made up of the "B," "C" and "No" accident groups, and suggests the possibility of a difference between the mean reaction time of this population and that of the "A" accident category. However, the possibility that such a relationship is in reality a function of age, to which reaction time is apparently related, precludes any generalization of this result. It also remains possible that the classification of accidents used is not correct for determining a relationship between reaction-time variability and accidents. Reaction-time variability may be related to a class of accidents not used in this study. Basic research is needed to establish representative classes of accidents.

Summary

The purpose of this study was to test the hypothesis that an individual whose reaction times are relatively consistent throughout a series of measurements, irrespective of the magnitude of his mean reaction time, will be found to be less susceptible to automobile accidents than is the individual whose reaction times are more variable.

A reactometer was employed in making simple reaction-time measurements on a sample of 104 Radio Patrol Officers. The reactometer measures the time required for the subject to move his foot from an "accelerator" pedal to a "brake" pedal in response to a small, red stimulus light. Ten successive measurements were recorded for each subject.

A mean and a standard deviation were computed for each subject's distribution of 10 reaction times. The mean reaction times ranged from .348 to .563 seconds with a mean of .433 seconds. The standard deviations ranged from .0156 to .0983 seconds with a mean of .0393 seconds. The statistical reliability of the means calculated by the split half technique was .87. Correction for the number of meas-

ures used yielded a reliability coefficient of .93. The uncorrected reliability of the standard deviations was .27 which similarly corrected yielded a reliability coefficient of .41.

The basic criterion data consisted of the official reports of accidents sustained by the subjects while on duty as drivers in police patrol cars. These accidents were subjectively classified into three different categories as follows:

"A" accident—an accident in which the officer was apparently at fault; an "avoidable" accident.

"B" accident—an accident in which responsibility was doubtful or indeterminable.

"C" accident—an accident in which the officer was apparently not at fault; an "unavoidable" accident.

An initial analysis of the data was made by means of a tetrachoric correlation technique.

A statistically significant correlation (1 per cent level of confidence) was found between mean reaction time and age.

A statistically significant negative correlation (5 per cent level of confidence) was found between mean reaction time and accidents when accidents were "split" on the basis "B," "C" and "No" accidents *vs.* "A" accidents. The correlation, however, was nonsignificant at the 5 per cent level when age was partialed out.

No statistically significant correlations were found between standard deviations of reaction time and (1) age, or (2) mean reaction time, or (3) accidents.

Using an analysis of variance technique, it was found that there was no statistically significant difference between the mean reaction time of accident groups "B," "C" and "No."

Using a modified *t*-test, it was found that there was a statistically significant difference (5 per cent level of confidence) between the mean reaction time of accident group "A" and the mean reaction time of accident groups "B," "C" or "No" when the latter were treated as a single group. However, here the age effect was not taken into account.

The results of this study do not substantiate the hypothesis that variability of reaction time is positively correlated with susceptibility to automobile accidents under the conditions of the present study. Certain basic research is needed to establish the generality of these results.

Received May 19, 1950.

References

1. Cobb, P. W. *Automobile driver tests administered to 3663 persons in Connecticut, 1936–37, and the relation of the test scores to the accidents sustained.* Unpubl. report to Highway Research Board, Washington, D. C., July 1939.
2. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans, Green & Company, 1947.
3. Johnson, H. M. The detection and treatment of accident-prone drivers. *Psychol. Bull.*, 1946, **43**, 489–532.
4. Mintz, A., and Blum, M. L. A re-examination of the accident-proneness concept. *J. appl. Psychol.*, 1949, **33**, 195–211.
5. Snedecor, G. W. *Statistical methods applied to experiments in agriculture and biology.* Ames, Iowa: Collegiate Press, 1938.

Psychological Climate and Accidents in an Automotive Plant *

Vernon Keenan and Willard Kerr
Illinois Institute of Technology
and
William Sherman
International Harvester Company—Tractor Works

In the belief that certain factors of psychological climate and physical environment may be of great importance in the causation of accidents in heavy industry, the authors formulated a series of hypotheses concerning correlates of accidents and subjected each to an experimental design for testing. The hypotheses were:

1. Excessive variation in pressure to "get out production" induces stress and tension contributing to accidents.
2. Intensity of normal production schedules contribute to accidents.
3. High promotion probability encourages alertness which reduces accidents.
4. A comfortable shop environment reduces accidents by minimizing physical annoyances and distractions.
5. Emphasis on individual responsibility (rather than upon crew work) acts to reduce accidents.
6. A reasonable level of job prestige is required for safe behavior.
7. Heavy manual work induces stress, fatigue, and weariness, which increase probability of accident.
8. Incentive pay systems encourage alertness which in turn promotes safe behavior.
9. Relative adequacy-predictability of work space is essential for avoidance of distraction-worry and safety.
10. Presence of an obvious and threatening danger has a semi-hypnotic effect upon the worker and increases accidents which are unrelated to the obvious danger.

Subjects

Lost-time accident records of 7,103 (average) personnel of a tractor factory for the 1944–48 five-year period were tabulated for each of the 44 shop departments of the plant, and accident rate per 100 personnel was computed for each department. The enterprise experienced 1,943 lost-time accidents during this period, and departmental rates per hundred personnel per year ranged from 1.5 to 13.2.

Procedure

Five-point continuum ratings on each of the variables shown in Table 1 were collected independently from each member of a "best-informed" panel of judges. This panel of eight judges, all individuals with at least ten years of diversified experience in the plant, includes the staff assistant to the works manager, the assistant works manager, the general superintendent, the plant engineer, the chief of methods and rates, the planning engineer, the safety supervisor, and the industrial relations manager. Split-half (odd-versus-even judges) reliability coefficients corrected via the Spearman-Brown prophecy formula for these ratings are indicated in Table 1.

Mean judge ratings on each department, on each of the ten variables, then were computed and the departmental means on each variable were correlated with departmental accident rates. Intercorrelations among all variables also are shown in Table 1.

Results

Results of these correlational studies between accident rates and certain psychological characteristics of organizational units are displayed in Table 1. A summary inspection of the table suggests that several of the ten hypotheses stated receive enough confirmation to continue them as tenable into more rigorous future researches.

* The authors gratefully acknowledge the indispensable aid and original suggestions of Mr. Howard M. Huntington, Safety Supervisor, International Harvester Tractor Works.

Variation and Volume of Production Pressure. Neither of these variables seem to be significantly related to accident rate. The extraordinary fact of a negative (−.42) reliability cofficient suggests the chance identification of two "schools of thought" as to what constitutes "pressure to get things done." Since each of these variables correlate substantially with at least one other variable in Table 1, it is obvious that the split-half reliability coefficients in these two instances are gross underestimates.

Partial correlations of $r_{1,11\cdot9} = -.18$ and $r_{1,11\cdot4} = -.55$ help estimate possible effects of incentive work and comfort of shop environment on the relationship between accidents and rated variation in production pressure. When comfort of shop environment is held constant by partial correlation, the relationship increases to statistical significance, suggesting *fewer* accidents in departments with above average variation in production pressure. While all statistics in this study which pertain to production pressure are somewhat questionable in interpretation because of the apparently divided opinions of raters, it is possible that variation in production pressure tends to reduce accidents by reducing monotony and encouraging alertness.

Similarly, when comfort of shop environment is held constant, the partial correlation between constant production pressure and departmental accident rate attains (inverse) a value of unity. If this result is not artifactual, it may be that production pressure even when constant encourages alertness which in turn makes for safe behavior.

Promotion Probability. Although the zero order correlation between departmental promotion probability and accident rate is insignificant, the relationship becomes significant (−.90) when comfort of shop environment is held constant by partial correlation. This result is consistent with the finding of a significant correlation of −.40 between these same two variables in a New Jersey electronics plant.[1]

Comfortable Shop Environment. Most significant of the findings in this research is the discovery of a correlation of −.70 between accident

[1] Kerr, W. A. Accident proneness of factory departments. *J. appl. Psychol.*, 1950, **34**, 167–170.

Table 1

Pearsonian Intercorrelations among Accident Rate and Each of Ten Other Variables in 44 Departments of a Tractor Factory (7,103 Personnel) *

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Variation in Pressure to Get Things Done	*−.42*	.27	.00	**−.60**	**.56**	**−.30**	.24	**.55**	**.78**	.11	.11
2. Constant Pressure to Get Things Done	.27	*.26*	.23	**−.94**	**.42**	−.08	**.84**	**.67**	**.50**	**.65**	−.06
3. Promotion Probability	.00	.23	*.92*	**−.67**	−.28	**.94**	−.11	.00	**.56**	**.43**	−.01
4. Comfortable Shop Environment	**−.60**	**−.94**	**−.67**	*.82*	**−.50**	**.44**	**−.46**	**−.59**	−.11	**−.53**	**−.70**
5. Degree of Crew Work	**.56**	**.42**	−.28	**−.50**	*.91*	.00	**.58**	**.46**	.13	**.37**	**.33**
6. Job Prestige	**−.30**	−.08	**.94**	**.44**	.00	*.99*	**−.30**	**−.34**	.05	.18	−.23
7. Manual Effort Involved	.24	**.84**	−.11	**−.46**	**.58**	**−.30**	*.83*	**.68**	**.36**	**.64**	**.47**
8. Degree of Operational Congestion	**.55**	**.67**	.00	**−.59**	**.46**	**−.34**	**.68**	*.70*	**.49**	**.47**	**.45**
9. Degree of Incentive Work	**.78**	**.50**	**.56**	−.11	.13	.05	**.36**	**.49**	*.76*	**.43**	.28
10. Degree of Obvious Danger Factor	.11	**.65**	**.43**	**−.53**	**.37**	.18	**.64**	**.47**	**.43**	*.95*	**.35**
11. Accident Rate	.11	−.06	−.01	**−.70**	**.33**	−.23	**.47**	**.45**	.28	**.35**	*.73*

* Coefficients in boldface type are statistically significant at the five per cent level of confidence or better. In determining the statistical significance, the formula $\sigma_r = \frac{1 - r^2}{\sqrt{N - 1}}$ was used. Coefficients in italics represent odd-even–corrected reliability coefficients, except for accident rate which is not corrected.

rate and mean-rated comfort of the shop environment. Partial correlations holding constant degree of operational congestion (−.60), manual effort involved (−.55), constant pressure (−.96), and degree of crew work (−.66) sustain the original correlation. The psychological implication of this result is that approximately half of the variance in departmental accident rates is accounted for by variance in the comfort factor. Heavy industries in which forges and foundries are essential operations inevitably experience great difficulty in making some job environment physically comfortable. Heat radiating from molten metals, violent sound vibrating from drop forges, and gritty earth particles from foundry "molds" are vivid examples of the practical obstacles to comfortable adjustment of the worker to the job.

Workers with low "physical frustration tolerance" under conditions of heat, noise, etc., probably have somewhat higher accident frequency than their co-workers. An accumulation of physical frustration tensions leads easily to "distractive" behavior which often results simultaneously in reduced quality of work and accident. It should be emphasized, however—and this qualification is supported by the evidence of Mintz and Blum[2]—that variance in such accident proneness of indivudals probably accounts for a minority of the variance in accident rates. In other words, the "normal" worker as well as the "prone" worker is prone to these same unsafe behaviors under identical frustrating conditions, but at a slightly better threshold of tolerance. A reasonable conclusion under such evidence appears to be that *major preventative effort can most profitably be expended in working to alleviate as much as possible the overall physical frustrations which impinge on all personnel.* Such effort can reasonably be expected to reduce the accident rates of both the more-prone and less-prone personnel.

Degree of Crew Work. Departments in which the typical job performance is in a "crew" often have above-average accident rates. Crew work in difficult operations (e.g., moving a hot ingot to position in a drop forge and later removal of the shaped part calls not only for reasonably synchronous action and cooperation of the crew members, but constant vigilance for the continous accepting and relinquishing of responsibility as well as the development of any unusual circumstances. Less stringent cooperative work sometimes has greater accident probability also in large part perhaps because of the occasional transition moments in which *responsibility* for consequences is vague or suspended but operations nevertheless continue, from momentum if nothing else. In those brief seconds, accidents are made. In individual-type work the individual rarely is uncertain about responsibility for consequences.

Job Prestige. It is interesting to note that while this correlation of −.23 with accidents is not statistically significant, it is highly consistent with the value of −.30 found in the 53 departments of the Camden Works of RCA.[3] Combining the two plants makes a total of 97 departments with a total of 18,000 personnel, and the relationship between job prestige and accident rate is established as highly significant. Partial correlation evidence, however, suggests that the relationship is at least in part artifactual. When comfort of shop environment is held constant, the coefficient declines to .12, and when manual effort is held constant it declines to −.10. High accident departments are characterized by less physical comfort and a demand for greater manual effort; incidentally, they are of lower average prestige value, but this fact seems to have little causal bearing upon accident rate.

Manual Effort. The greater the degree of manual effort involved in successful performance of the typical job, the greater the tendency toward accidents ($r = .47$). This relationship persists ($r_{7,10\cdot2} = .96$) when "pressure to get things done" is held constant by partial correlation. As in the demand for constant vigilance in crew work, the demand for physical energy sometimes exceeds efficiently directed supply. Workers trained in economical expenditure of effort as well as in control of physical forces may reduce the frequency of such accidents.

Operational Congestion. "Degree of operational congestion" implies frequency-proximity of moving objects and persons. Such activities in the psychological periphery of the immediate

[2] Mintz, A., and Blum, M. L. A re-examination of the accident-proneness concept. *J. appl. Psychol.*, 1949, 33, 195–211.

[3] Kerr, W. A. *Op. cit.*

task may connote greater annoyance-distraction value than actual physical threat to the worker, but may contribute ($r = .45$) to unsafe behavior of personnel. A limitation to this interpretation is the fact that the correlation declines to .06 when comfort of shop environment is held constant by partial correlation. In view of this limitation, a causal interpretation of the zero order r is not warranted.

Incentive Work. Departments with incentive work may on the average incur a few more accidents than do departments on the "straight-time" payment system. This is suggested by the original correlation of .28 (not significant) which is increased to .35 when promotion probability is held constant. While evidence is not available to test the point, it is possible that when more work is being performed (as it usually is under an incentive system) more accidents will occur. Thus, an increase in accidents under incentive conditions might be a function of increased productivity or productive activity ("motionless bodies" rarely have accidents). On the other hand, this finding barely borders on statistical significance and similar study of another factory yielded a correlation of .00 between incentive system and departmental accident rate.[4]

Obvious Danger Factor. Departments with a constant obvious danger present such as glowing molten metal have significantly higher accident rates ($r = .35$) than do other departments. An arresting finding by Mr. Huntington, the Safety Supervisor, on the nature of the accidents in these departments is that the accidents usually are not physically involved with the obvious hazard factor. For example, in molten metal operation departments, accident rates are above average but they rarely include burns. The obvious hazard seems to exercise an almost hypnotic effect upon personnel, delimiting their attention too much and facilitating their involvement in other hazards. When comfort of shop environment is held constant (this is a questionable procedure in this instance because "comfort" and "danger factor" may be regarded as almost synonomous in certain key departments), the zero order r declines to a partial r of $-.04$ between obvious danger factor and departmental accident rate.

[4] Kerr, W. A. *Op. cit.*

Summary

Against a criterion of five years of lost-time accident rates in the 44 shop departments of a tractor factory, ten possible explanatory hypotheses were evaluated by means of data from a special rating panel. Until more exacting and precise data are obtained, all conclusions of this research must be considered tentative; however, within these limitations, certain trends appear.

1. Results obtained on variation in production pressure and constant production pressure are inconclusive.

2. The hypothesis that promotion probability encourages safe behavior appears to be supported by this study.

3. Comfortable shop environment appears from this research to be a major determinant of safe behavior.

4. Degree of crew work is positively correlated with accident rates of the factory departments in this study. This is believed to be the result of more complete and continuous acceptance of individual responsibility in the non-cooperative work situation.

5. Job prestige of typical job fails to correlate significantly with departmental accidents when certain other variables are held constant.

6. The greater the degree of manual effort involved, the higher, on the average, is the departmental accident rate.

7. Both degree of operational congestion and degree of incentive work fail to predict departmental accident rates when certain other variables are held constant.

8. The effects of an "obvious danger factor" on departmental accident rates are somewhat uncertain, although observational evidence indicates that where an obvious danger factor exists, the accidents which occur tend *not* to be identified with the obvious danger. Existence of an impressive obvious hazard seems to contribute to "unrelated" accidents by delimiting attention and encouraging proneness to involvement in the non-obvious hazards.

Received June 5, 1950.

Industrial Use of the Partington Pathways Test

Lee W. Cozan

Hechinger Co., Washington, D. C.

No general claims of superiority regarding reliability and validity may be made in behalf of the short mental ability test in comparison with the longer test. However, Otis and Chesler (**6**) have pointed out several advantages in the utilization of the short mental ability test in the industrial situation. In the case where time is at a premium, the short test of mental ability may be used with more than acceptable validity. Where the selection standards are more exacting, the short test may be used as a screen to indicate those applicants who are definitely below or above the arbitrarily set mental level. This is a time-saving and economical procedure inasmuch as these applicants are not administered the entire battery of tests.

A survey of eight paper-and-pencil tests of mental ability (**1**, **2**, **4**, **5**, **9**, **10**, **11**, and **12**) suitable for employment purposes by business and industry, most of which are listed in the Third Mental Measurements Yearbook (**3**), revealed the "working time" ranged from 10 to 50 minutes with the total administration time varying from 12 to 60 minutes. The means for "working time" and administration were 24 and 30 minutes, respectively. See Table 1. To all appearances, there are a few short tests of mental ability suitable for use in the industrial situation. A short rest is arbitrarily defined as 15 minutes or less. Only three of the eight tests required 16 minutes or less.

Table 1

Distribution of Time for Eight Mental Ability Tests

Test	Working Time	Administration Time
Adaptability	15	20
Classification Test for Industrial and Office Personnel	10	12
Benge Employment	22	30
Bureau Test VI—Mental Alertness*	15	25
Otis Self-Administering	30	35
Scovill Classification	50	60
Wesman Personnel Classification	28	35
Wonderlic Personnel	12	20

* This test is no longer published.

Prompted by this information, a research project was undertaken to determine the suitability of the Partington Pathways Test (**7**) as a short test of mental ability in the industrial situation.

Description of the Test

The Partington Pathways Test, initially called the Distributed Attention Test, was originally devised in 1938 as a test of speed of performance. However, it appeared that successful performance on the test demanded the integration of previously learned material into a complex pattern of response requiring shifts and fluctuations in attention, accuracy and speed, organization, retention, recall, and synthesis.

When the test was administered in 1942 to an unselected adult male population in order to develop a standardized procedure for administering the test and a method of scoring, the results led the authors to draw the conclusion that the test was a measurement of general mental ability because the scores on the test were in general agreement with the estimated intelligence of the individuals to whom the test was administered. The conclusion was substantiated when the test was administered to a random sampling of a group of 256 male World War II veterans, ranging in age from 19 to 36, who had been given the 1937 Revision of the Stanford-Binet Scale of Intelligence. The coefficient of correlation between the scores on the Pathways Test and the Revised Stanford-Binet Scale was found to be 0.68 (**8**).

The Pathways Test has two forms. In Form I, the task requires the subject to draw a line connecting numbers from 1 to 25 in order beginning with number 1. The numbers are arranged in random order. Form II uses both numbers and letters and requires a line be drawn commencing with number 1 to letter A,

Table 2

Average and Variability of Scores on the Otis S-A Test and Pathways Test

	Otis S-A	Pathways
Range	20–66	6–40
Mean	41.5	21.6
Standard Deviation	11.7	8.2

from letter A to number 2, then from number 2 to letter B, and so on. The time in seconds required to finish each pathway is the score. The time required is converted to a raw score by means of two tables (8).

Results

For the purpose of experimentally applying the Partington Pathways Tests in an industrial situation, a group of 80 clerical and skilled employees selected at random were administered the test. The group also had been given given the Otis Self-Administering Test of Mental Ability.[1] The correlation coefficient between the scores made on the two tests was found to be 0.74. Table 2 summarizes the information about the range, mean, and standard deviation of the subjects on the Pathways Test and Otis Test of Mental Ability.

The average "working time" on the Pathways Test was 135 seconds, ranging from 78 to 255 seconds. Administration time ranged from 5 to 9 minutes, averaging approximately 7 minutes.

Summary

A short test of mental ability has been presented which appears very practical for use in the industrial situation. Its chief advantages are: (1) a short "working time"; and (2) ease of administration and scoring. Its one drawback is that being a speed test it is disadvantageous to use with older applicants since it has been found that older adults do not perform so well on tests depending chiefly upon speed of response. At the present writing, the test has been validated with another test, the Otis Self-Administering Test, with a resulting correlation coefficient of 0.74.

Received June 29, 1950.

[1] A special printing of the test without the words "mental ability" was used.

References

1. Anderson, J. E., and Pond, M. *Scovill classification test.* New York: Psychological Corporation.
2. Benge, E. J. *Benge employment test.* New York: Management Service Company, 1942.
3. Buros, O. K. *The third mental measurements yearbook.* New Brunswick, N. J.: Rutgers University Press, 1949.
4. Otis, J. L., et al. *Classification test for industrial and office personnel.* Cleveland, Ohio: Western Reserve University Press, 1947.
5. Otis, A. S. *Otis self-administering test of mental ability.* Yonkers: World Book Company.
6. Otis, J. L., and Chesler, D. J. A short test of mental ability. *J. appl. Psychol.*, 1949, **33**, 146–150.
7. Partington, J. E., and Leiter, R. G. *Partington pathways test.* Washington, D. C.: Psychological Service Center, 1949.
8. Partington, J. E., and Leiter, R. G. Partington's pathways test. *The psychological service center bulletin*, 1949, **1**, 9–20.
9. Tiffin, J., and Lawshe, C. *Adaptability test.* Chicago: Science Research Associates, 1942.
10. Wesman, A. *Wesman personnel classification test.* New York: Psychological Corporation, 1947.
11. Wonderlic, E. J. *Personnel test.* New York: Psychological Corporation.
12. Yoakum, C. S. *Bureau test VI—mental alertness.* New York: Psychological Corporation, 1920.

Mirror-Tracing Practice for Medical and Dental Students

Norma V. Scheidemann

Los Angeles, California

Dentists and nose-throat specialists conduct a considerable proportion of their professional practice in mirror-vision, with the mirror placed at every possible angle in relation to the arena of operation. General medical practitioners have a limited use for mirror-performance since they employ it with the indirect method for removing objects lodged in a mouth passage.

Students of medicine and dentistry routinely develop mirror-performance skill incidentally, during laboratory and clinical practice, by more or less haphazard methods with attention focused upon professional problems.

It is suggested that these students develop mirror-performance skill through isolated systematic mirror-tracing practice with attention focused upon orientation in mirror-spaces presented in a mirror placed in a wide variety of positions, before clinical practice involving mirror-vision is undertaken.

Several items are available[1] for conducting such practice expeditiously, namely:

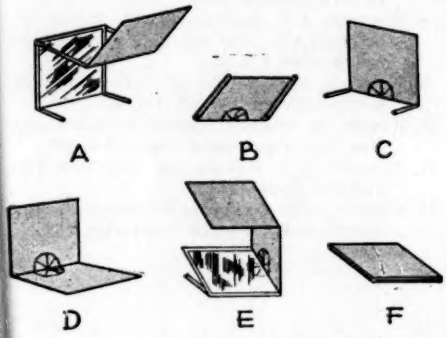

FIG. 1. A mirror-scope. *A*, Mirror-screen with supports; *B*, metal square with conduits; *C*, metal square with supports; *D*, performance planes (supports of *C* inserted into conduits of *B*); *E*, assembled mirror-scope (supports of *A* and *C* inserted into conduits of *B*); *F*, collapsed mirror-scope. *B* and *C* have compass markings for gauging direction of tracing on performance planes.

[1] Available from the C. H. Stoelting Company, Chicago, Illinois.

A Mirror-Scope (Fig. 1). This device provides for: (1) the reference mirror-spaces—forward, backward, leftward, rightward, upward, and downward—and all mirror-spaces at any angle to any of them; (2) mirror-vision in any mirror-space; (3) three-dimensional performance planes for any mirror-space; and (4) compass markings on performance planes for gauging direction of tracing at any angle, on any of the three-dimensional planes, in any mirror-space.

The Compass Diagnostic Test of Directional Mirror-Tracing Skill (Fig. 2). This test provides, for diagnostic purposes, objective records of mirror-tracing skill in 24 radial directions on each of the three-dimensional planes, in each of the six axial mirror-spaces.

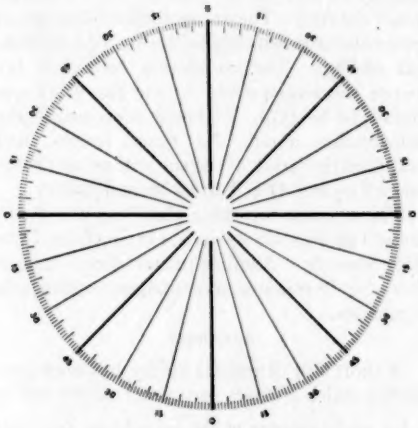

FIG. 2. Compass test-pattern.

Trace-and-Score Mirror-Tracing Practice-Patterns. These are three patterns of graded difficulty printed on a scoring scale. *A Linear Practice-Pattern* (Fig. 3) consists of 10 parallel lines that can be gauged to direct tracing at any angle on the three-dimensional planes, in any mirror-space. *A 100 Angles Practice-Pattern*

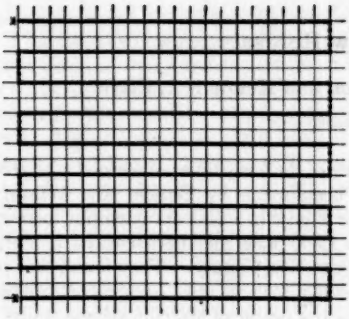

FIG. 3. Linear practice-pattern.

(Fig. 4) consists of a continuous line featuring 100 right angles. It is printed in two sizes. The large size directs practice in directional change in tracing axial directions in any mirror-space. The small size has shorter lines, which offer less opportunity for directional orientation between directional changes. This size can be gauged for practice in directional change in tracing axial lines and also lines at any specific angle on any of the three-dimensional planes in any mirror-space.

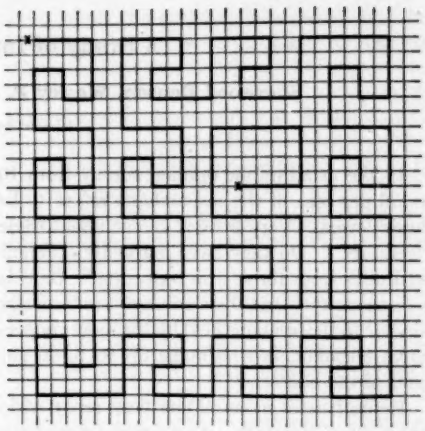

FIG. 4. 100 angles practice-pattern.

Received January 5, 1951.
Early publication.

A Study of Interview Refusals

Sherwood Benson, Wesley P. Booman, and Kenneth E. Clark *

University of Minnesota

In the usual public opinion survey from 10 to 20 per cent of the respondents refuse to be interviewed. Very little is known about these "refusals." Do they have markedly different attitudes from those who cooperate? Or is it safe to assume that their opinions are basically similar to the opinions held by the more cooperative respondents who make up 80 to 90 per cent of the universe?

Refusal rates have seldom been reported in the polling literature. Some researchers have regarded the reported refusal rate of 10 to 20 per cent as alarmingly high (**3**). Refusal rates may be even higher—up to 30 per cent (**5**). Others express little concern (**1, 6**). They assume that refusals are not likely to differ in important respects from those who respond. However, neither group has published evidence supporting their conclusions. Certainly such evidence is essential before we can conclude that failure to complete refused interviews is causing a biased result.

In this study we were concerned with two problems: (1) How can the refusal rate be cut down to insignificance? and (2) how do the refusals, if successfully completed, affect the survey results? Heneman and Paterson (**3**, p. 393) report that for many characteristics, persons who refuse to cooperate have decidedly different characteristics from those who do cooperate. However, the study which they cite (**4**) deals not with refusals but with respondents who were initially not at home. Furthermore it is not enough to know that those who refuse, or are not at home, have decidedly different characteristics from those who can be contacted. Unless such characteristics are related to the survey results, different characteristics *per se*, do not assure the pollster that the exclusion of the "refusal" or "not at home" is introducing any bias.

One of the few studies concerned with refusals has been made by Gaudet and Wilson (**2**). Their study is based on attempts to re-interview 1,800 panel members selected as representative of Erie County. They found that refusals tended to come most frequently from housewives and people of low educational status. It should be noted, however, that their study dealt not with original refusals but with people who refused to be re-interviewed.

Procedure

The questionnaire used in this study was constructed by 60 members of a public opinion analysis class at the University of Minnesota and dealt with attitudes toward Negroes. It included twelve questions of the sort, "If you were in a hospital, would you mind having a Negro nurse?" A fixed address sample was selected by the instructor by listing every separate dwelling unit in a class C socio-economic residential area in South East Minneapolis.

Students enrolled in the public opinion class conducted the initial interviews. This group, which might be expected to have a better than average grasp of the problems of the public opinion survey, though they were somewhat lacking in experience, probably compares satisfactorily with the usual survey interviewers. With instructions to complete an interview with a "responsible adult" living in the unit, and to make up to three calls at a fixed address in the event the respondent was not at home, this group completed 173, or 75.2 per cent, of 230 assigned interviews. Twenty-four, or 10.4 per cent, were still not at home after three calls, and 33, or 14.4 per cent, refused to answer the questionnaire.

The reasons for refusals as reported by the original interviewers were: Very busy, 11 or 33 per cent; Sickness, 7 or 21 per cent; Not interested, 2 or 6 per cent; and No reason, 13 or 39 per cent.

The problem facing the writers then was, first, to contact all respondents reported not at

* The planning of the interviews of refusals, and the actual interviewing, was done by the first two of the three writers as part of their work towards the Master of Arts degree at the University of Minnesota.

home, and second, to attempt to complete all the interviews which were initially refused. After making as many as five additional interviews the 24 not-at-home respondents were successfully interviewed. Most of the interviewees (both refusals and not-at-homes) were ultimately contacted between 6:30 and 7:30 p. m. The calls required to complete these 24 and the original 173 are shown in Table 1.

After most of the not-at-home respondents were contacted a method for completing the refused interviews was developed. Purely through armchair stratagem it was decided that the best results in this particular study might be obtained by using a young and pretty female interviewer.[1] One thing was clearly evident from the record of the refusals: most people refused before they even knew what the questions were about. In training the interviewer strong emphasis was placed upon her appearance and a friendly manner.

Numerous test interviews were conducted with the interviewer, the writers playing the role of reluctant respondents. A variety of different answers, based upon the anticipated objections of respondents, were prepared and rehearsed. Since practically all of the refusals occurred before any survey questions were asked, such things as a friendly smile and an easy and confident greeting were stressed. It was also suggested that the interviewer: (1) keep talking as long as possible even though it appears certain that the respondent will refuse; and (2) not step back or retreat from the doorway until the respondent actually closes the door. By accompanying the interviewer and observing her methods from a car the writers were able to offer suggestions and criticism as the survey progressed.

On her first call our female interviewer completed 29 of 33 attempted interviews with respondents who had initially refused to cooperate. In three of the four remaining cases, the writers, believing that less resistance might be encountered by another interviewer, attempted to re-interview the respondents. Before any re-interviewing, however, follow-up letters describing the survey and asking for cooperation were sent to the four remaining refusals. After the letters were sent the female interviewer successfully completed one and the writers two out of three of the re-interviews. When the remaining respondent had refused to cooperate it was decided to send him a second letter which included a questionnaire and a stamped, self-addressed envelope. No reply was received.

[1] Approximately three out of four of the original interviewers were men.

Table 1

Summary of Completed, Not-Home, and Refused Interviews Arranged by Call and Showing Time Required to Complete Each Set of Interviews

	Completed	Cum. % Time to Complete*	Not-Home	Refused	Yet-to-Complete
1st Call	136	50	70	24	94
2nd Call	29	66	40	1	65
3rd Call	8	75	28	4	57
4th Call	15	82	10	3	42
5th Call	5	85	4	1	37
6th Call	2	86	2	0	35
7th Call	1	87	1	0	34
8th Call	1	88	0	0	33
Additional Calls on 33 Refusals:					
1st Call	21	95	9	3	12
2nd Call	8	98	3	1	4
3rd Call	2	99	0	2	2
4th Call	1	100	0	1	1

* Per cent of total interviewing time required to complete 229 of the 230 interviews, assuming an interviewer using an automobile, and living 5½ minutes from the interviewing area.

Results

By carefully training and selecting interviewers it was possible to complete 99.6 per cent of the interviews in a fixed address sample. (N = 230.) Since the refusals occurred before any questions were asked, appearance and manner of the interviewer seemed to be a major factor in reducing refusals.

In order to complete interviews originally refused, about 14 per cent additional time was required. By trying both to contact the not-at-home respondents after three calls and to complete the refusals, about 33 per cent more time was required. This does not include the additional time necessary to treat these respondents on a more individual basis, which consisted of planning interviews, training the interviewer, and writing letters.

Table 2

Comparison of Refusals and Non-Refusals on Four Variables

	Attitude toward Negroes	Years Education	Age	% Union Members
Means:				
Total Group (N = 229)	15.83	11.21	43.4	46.7
Non-Refusals (N = 197)	16.22	11.43	43.3	46.7
Refusals (N = 32)	13.41	9.88	43.6	46.9
Mean Differences, Refusals—Non-Refusals	2.81	1.55	0.3	0.2
SE_{diff}	0.959	0.595	—†	—†
t	2.93*	2.61*	—†	—†

* Significant at the one per cent level.
† Non-significant, not computed.

A comparison of the original refusals and the non-refusals was made on four variables: mean total score on attitudes toward Negroes,[2] years of education, age and union membership. These comparisons are summarized in Table 2.

[2] Questionnaires were scored by assigning a value of *2* to each favorable response, a *1* to an equivocal response, and a *0* to an unfavorable response. For the twelve items, scores ranged from 24 (perfect favorable score) to 0.

Both in attitudes towards Negroes, and in years of eduaction, those respondents who initially refused to be interviewed differ significantly from those who did not, the refusals being less favorable towards Negroes, and having less education. No appreciable difference was found between the two groups in either age or union membership.

Since the refusals constituted only 14.4 per cent of the sample, their responses, when combined with those of non-refusals, did not affect either the total score on attitudes toward Negroes or the mean education level to any appreciable extent. This was true also in responses to individual items of the questionnaire, in that no percentage response would have deviated more than 3.4 points had the refusals been omitted from the study. (Even so, these differences are larger than the differences which would have obtained if the not-at-homes after three calls were withdrawn from the sample.) Had our refusal group been larger, as is probably the case for many surveys, the net effect of the sort of difference observed here would have been magnified.

Discussion

In this study some indication has been given of the amount of time required to include in one's sample the refusals or those not-at-homes requiring more than three calls to contact. About 14 per cent more time was required for interviewing respondents in order to include responses from those who initially refused to cooperate. About 33 per cent more time was required to complete interviews both with this refusal group and with those not at home for the first three calls.

It is apparent that the decision on how much effort should be expended to complete all interviews is a function of the particular type of question asked and the accuracy desired. If it is suspected that significantly different answers would be obtained from refusals as compared with non-refusals (as was found in this study), it would be wise to make a concentrated effort to complete enough refusal interviews to establish trends.

On the basis of the experience gained in this study, the writers would recommend:

1. That any agency seriously concerned with survey methods train its interviewers so as to reduce initial refusals,

2. That interviewers record on the schedules for refused interviews a short description of the refusing respondent, for later identification to attempt to complete the interview, and some indication of the reason for refusal, and

3. That refusals be re-interviewed in any survey in which the number of refusals exceeds some minimum figure (say, 5 or 10 per cent), to determine whether this non-contacted group may influence to any appreciable degree the results of the survey.

Received June 9, 1950.

References

1. Gallup, G. *A guide to public opinion polls.* Princeton: Princeton University Press, 1948.
2. Gaudet, Hazel, and Wilson, E. C. Who escapes the personal investigator? *J. appl. Psychol.*, 1940, **24**, 773–777.
3. Heneman, H. G., Jr., and Paterson, D. G. Refusal rates and interviewer quality. *Int. J. Opin. Attitude Res.*, 1949, **3**, 392–398.
4. Hilgard, E. R., and Payne, S. L. Those not at home: Riddle for pollsters. *Publ. Opin. Quart.*, 1944, **8**, 245–261.
5. Parry, H. J. Impact of the Kinsey Report on opinion and attitude research. *Sci. Mon.*, 1950, **70**, 286–290.
6. Sheatsley, P. B. Public relations of the polls. *Int. J. Opin. Attitude Res.*, 1948, **2**, 461–465.

Social Characteristics of Respondents and Non-Respondents in a Questionnaire Study of Later Maturity

Ju-Shu Pan

Chicago, Illinois

The problem of selective response in questionnaire studies of adjustment of persons 65 and over is important because of the increased chance of bias in favor of the better adjusted group. Non-response bias was analyzed in a study of personal and social adjustment of persons institutionalized in homes for the aged in the United States.

Two hundred and sixty-three residents of 26 homes sponsored by religious organizations filled out the schedule "Your Activities and Attitudes." These homes had a total of 2,566 residents. Nine of the homes were sponsored by the Methodist Church and 17 were controlled by other Protestant groups.

Schedules were distributed through superintendents of the homes. Some refused to distribute schedules in their institutions, often stating the residents were "too old." Others selected certain residents as respondents and in some cases asked for volunteer respondents.

Superintendents of the same 26 homes were asked to prepare one page confidential "check lists" on every tenth resident, ranked alphabetically. Check lists provided data on sex, age, education, occupation, health, and number of physical problems. Check lists for 335 (13 per cent) of the 2,566 residents were received.

Data for the randomly selected check list respondents were compared with the biased data for the 263 schedule respondents. The schedule sample was expected to be younger, better educated, and in better health than the check list sample. Non-respondents were expected to be those residents suffering the greatest impairment of physical and psychological function.

Mean age of the 263 respondents was 76.4 years with standard deviation of 9.3 years. The check list sample was slightly older, with mean age of 79.5 years and standard deviation of 11.4 years. Critical ratio is 3.3. Signifi-

Table 1

Proportions of Schedule and Check List Respondents by Selected Social Characteristics*

Social Characteristics	Schedule Respondents	Check List Respondents	Critical Ratio
Native-born	89.4	82.3	2.2
Education			
None	1.3	7.2	3.1
Grade	19.3	42.2	4.4
College	24.4	11.2	3.9
Professional	20.1	7.6	3.7
Former Occupation			
Minister	15.3	4.0	3.6
Clerical	26.8	15.0	2.7
Housewife	19.8	37.0	3.8
Health Rating			
Poor	19.0	29.0	2.5
Excellent	18.3	10.3	2.6
Physical Problems			
Four or more	3.8	8.8	2.0

* Number of cases on which percentages were computed ranges from 157 to 335.

Table 2

Proportions of Check List Respondents in Cooperating and Non-Cooperating Homes by Selected Adjustment Characteristics*

Adjustment Characteristics	Cooperating Homes	Non-Cooperating Homes	Critical Ratio
Nervousness			
Extremely nervous	0.9	10.7	4.4
Very nervous	5.3	19.6	4.9
Somewhat nervous	51.3	41.9	2.2
Not nervous	42.5	27.8	3.0
Happiness			
Very happy	20.5	34.8	3.8
Average	32.8	20.8	3.2
Adjustment			
Very well adjusted	41.3	31.3	2.5

* Number of cases on which percentages were computed ranges from 232 to 279.

cant differences by nativity, education, occupation, health, and number of physical problems are presented in Table 1.

Differences between residents of homes with schedule respondents and homes without schedule respondents, as measured by superintendents' check lists, are presented in Table 2.

Respondents filling out schedules and respondents for whom check lists were returned by superintendents differ significantly in the following ways: schedule respondents are older and more are native-born; more of the respondents had college or professional training, and higher proportions were former ministers or clerical workers; respondents rate themselves healthier than the total group rated by superintendents; fewer of the respondents have as many as four physical problems.

Check list residents of homes with schedule respondents were rated as significantly less "nervous" than were check list residents of non-cooperating homes. A higher proportion were "very well adjusted" but significantly fewer were rated as "very happy."

Received June 6, 1950.

MMPI and Student Teaching

John U. Michaelis and Fred T. Tyler

University of California

The diagnostic and predictive value of personality tests are of basic importance in planning for the use of such measures in a teacher education program. The reliabilities of various sub-tests and the intercorrelations between sub-tests are required in determining diagnostic value. The relationship of various sub-tests to teaching success gives an indication of predictive value. With diagnostic and predictive value in mind, the usefulness of a particular test can be given effective preliminary appraisal.

The purpose of this article is to make such an evaluation of the Minnesota Multiphasic Personality Inventory (2). The following nine scales are utilized in this study: *hypochondriasis, depression, hysteria, psychopathic deviate, masculinity-femininity, paranoia, psychasthenia, schizophrenia* and *mania*.

The data in this article are based upon 56 women students enrolled in student teaching. The mean chronological age was 23.5 years, with a standard deviation of 1.57. The range in age was from 21 to 41, 51 of the students being between 21 and 27 years of age. The test was given immediately before entrance into student teaching. Only those cases with low L (lie) and F (validity) scores are included in this study.

Diagnostic Value of Sub-Tests

The diagnostic value of sub-tests has been determined by the use of a technique first proposed by Kelley (3) and more recently reported by Bennett and Doppelt (1). The technique involves the calculation of reliability coefficients and intercorrelations of the various sub-tests. From these data the diagnostic value of sub-tests can be determined by means of a nomograph developed by Bennett and Doppelt (1).

The reliabilities of the sub-tests were determined by means of the split-half method and stepped up by the Spearman-Brown prophecy formula. Table 1 gives the reliabilities for a random sample of 100 women students tested in the elementary educational psychology classes at the University of California.

Reliabilities are very low on only two scales, namely, *paranoia* and *mania*. With the exception of *masculinity-femininity* (.61), all of the others are over .70.

In order to determine the diagnostic value of the sub-tests the odd-even reliabilities have been used along with the intercorrelations of the sub-tests which are reported in Table 1.

The intercorrelations vary from −.09 to .83,

Table 1

Corrected Split-Half Reliabilities of Sub-Tests and Intercorrelations between Sub-Tests*

	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Hypochondriasis	**.81**	.50	.32	.41	.23	.72	.58	.56	.17
Depression	.50	**.76**	.16	.36	.26	.04	.67	.54	−.01
Hysteria	.32	.16	**.87**	.21	.28	.15	.04	.09	−.09
Psychopathic deviate	.41	.36	.21	**.71**	.14	.16	.36	.47	.47
Masculinity-femininity	.23	.26	.28	.14	**.61**	−.01	.23	.23	.04
Paranoia	.72	.04	.15	.16	−.01	**.21**	.12	.06	.00
Psychasthenia	.58	.67	.04	.36	.23	.12	**.89**	**.83**	.23
Schizophrenia	.56	.54	.09	.47	.23	.06	.83	.84	.37
Mania	.17	−.01	−.09	.47	.04	.00	.23	.37	**.47**

* For d.f. = 100, r must be .197 and .256 to be significant at the .05 and .01 levels, respectively. For d.f. = 55, r must be .265 and .345 to be significant at the .05 and .01 levels, respectively (4, p. 212). The reliability coefficients in boldface type are based on 100 cases; the other coefficients are based on 56 cases.

Table 2

Diagnostic Values of Pairs of Sub-Tests

Sub-Test	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Hypochondriasis	—	.20	.33	.22	.22	*	.24	.22	*
Depression			.37	.22	.22	*	.16	.18	*
Hysteria				.30	.24	*	.46	.42	*
Psychopathic deviate					*	*	.28	.21	*
Masculinity-femininity						*	.26	.22	*
Paranoia							*	*	*
Psychasthenia								.10	.22
Schizophrenia									*
Mania									—

* Mean reliability correlations below .70 are not shown in the nomograph for evaluating pairs of tests. Asterisks indicate that proportion of difference in excess of chance cannot be determined from the nomograph.

16 being significant at the .05 level and 13 at the .01 level. In general, the intercorrelations are low enough to suggest that the various sub-tests are measuring different features of personality. It is of interest to note the intercorrelations between the tests constituting the neurotic and psychotic triads. The neurotic triad includes hypochondriasis, depression and hysteria, the intercorrelations being .50, .32 and .16, suggesting that if the different tests are measuring neuroticism they are measuring different aspects of it, at least in this group of college women. In the main the same applies to the sub-tests constituting the psychotic triad (paranoia, psychasthenia and schizophrenia), the intercorrelations being .12, .06 and .83; the psychasthenia scores correlate .83 with the schizophrenia scores, this being the highest correlation obtained in this sample.

The proportions of the differences between sub-tests in excess of chance have been determined by means of the nomograph designed by Bennett and Doppelt (1). This procedure utilizes intercorrelations between sub-tests and their reliabilities to determine the differential power of the sub-tests. This technique "estimates the proportion of all differences for that pair of tests which will represent true differences in ability" (1). The proportions are shown in Table 2.

Segel (5) states that a value of .25 is the minimum required for useful differentiation. On this basis only seven pairs of sub-tests provide useful diagnostic information, two of these pairs being members of the neurotic triad. If these values are verified in larger samples of comparable students it would appear that this test offers limited possibilities for diagnosis from the various sub-tests in the type of sample used in this study.

Relationship to Success in Student Teaching

Success in student teaching was appraised by university supervisors who have been using the same rating form for several years. Correlations of the sub-tests with the supervisor's ratings are shown in Table 3.

None of the sub-tests correlates high enough with student teaching to offer possibilities of predicting success in student teaching with any reasonable degree of accuracy.

In order to analyze further the differential power of the sub-tests, the sample of 56 students was divided into three groups on the basis of

Table 3

Correlations between Sub-Tests and Ratings in Student Teaching

Sub-Tests	Correlations with Rating
Hypochondriasis	−.05
Depression	−.08
Hysteria	−.36
Psychopathic deviate	−.13
Masculinity-femininity	−.05
Paranoia	.14
Psychasthenia	.01
Schizophrenia	.09
Mania	−.06

Table 4

Significance of Mean Differences between High and Low Groups

Sub-Test	Mean of High Group	Mean of Low Group	*t*-Ratio*
Hypochondriasis	3.82	3.86	.04
Depression	19.41	20.21	.50
Hysteria	19.88	23.21	2.57
Psychopathic deviate	12.12	14.43	1.86
Masculinity-femininity	37.65	38.21	.36
Paranoia	10.29	8.86	1.85
Psychasthenia	9.12	9.29	.08
Schizophrenia	8.47	7.29	.65
Mania	14.12	16.57	1.52

* With 29 degrees of freedom *t* must be 2.045 to be significant at the .05 level, and 2.756 at the .01 level.

the supervisor's ratings, high (N = 17), average (N = 25), and low (N = 14). The *t*-ratios (4) between the means of the high and low groups are given in Table 4.

Only one of the sub-tests, *hysteria*, differentiates sharply between the most and least successful student teachers. The difference on this sub-test is significant at the .02 level. *Psychopathic deviate* and *paranoia* are significant at the .10 level and should be studied further in a larger sample.

Low differential power has also been reported by Stough (6) in a similar study carried out with male high school teachers. Stough, however, did report differences between high and low groups on *depression* and *schizophrenia* which were significant at the .01 level, and on *psychasthenia* at the .05 level.

Summary

1. The reliabilities of the sub-tests of the Minnesota Multiphasic Personality Inventory reported in this study are fairly high with the exception of *paranoia* and *mania*.

2. The intercorrelations between sub-tests are low, but only a few pairs of sub-tests have diagnostic value insofar as the present sample is concerned.

3. In general, the sub-tests have very little value for predicting success in student teaching. However, the fact that there were differences between high and low groups on *hysteria*, *psychopathic deviate*, and *psychasthenia* suggests that these be given additional study with other samples. Similar study should also be made of several new scales which are now available for this inventory.

Received June 5, 1950.

References

1. Bennett, G. K., and Doppelt, J. E. The evaluation of pairs of tests for guidance use. *Educ. psychol. Measmt*, 1948, **8**, 319–325.
2. Hathaway, S. R., and McKinley, J. C. *Multiphasic personality inventory*. New York: The Psychological Corporation, 1943.
3. Kelley, T. L. A new method for determining the significance of differences in intelligence and achievement test scores. *J. educ. Psychol.*, 1923, **14**, 321–333.
4. Lindquist, E. F. *Statistical analysis in educational research*. Boston: Houghton Mifflin Co., 1940.
5. Segel, D. *Differential diagnosis*. Baltimore: Warwick and York, 1934.
6. Stough, M. F. The intercorrelations of measures of personality and success of student teachers. Unpublished doctoral dissertation, Univ. of Calif., 1949.

Correction for "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the MMPI"

Orpha Maust Lough

Rutherford, New Jersey

In the *Journal of Applied Psychology*, 1947, **31**, 437–445, a study was reported in which the Minnesota Multiphasic Personality Inventory was given in a liberal arts college to 115 unmarried women students, 61 of whom were cadet nurses and 54 liberal arts students, and in a state teachers college to 185 unmarried women students, 111 of whom were enrolled in the Music Curriculum and 74 in the General Curriculum for elementary school teachers. For Table 2 on p. 440 of the 1947 issue should be substituted Table 2, revised, presented on p. 126 of this issue.

From the data given in Table 2, revised, on the criterion of a critical ratio of 2.6 or more for statistical significance, there are reliable differences between the cadet nurses and the students in the General Curriculum on the Interest (M_f) scale, and between the cadet nurses and the Music students on the Hypochondriasis (H_s), Depression (D), Interest (M_f), and Schizophrenia (S_c) scales. These data indicate that the students in the General Curriculum and the Music students are reliably more feminine in their interests than the cadet nurses; or, the latter are more masculine in their interests than the former. It also appears that on the basis of the Hypochondriasis (H_s) scale, the cadet nurses are significantly more psychologically mature, tending to respond to adult problems with more adequate insight and less disposed to seek sympathy or to develop undue anxiety over their health, than the Music students. The cadet nurses are also significantly more self-confident and extrovertive, less prone to bizarre and unusual thoughts or behavior, and have better morale than the Music students. These data confirm and substantiate the trends analyzed in the original report.

In the Summary of the original report, substitution should be made, as follows:

2. There are statistically significant differences (critical ratio of 2.6 or more) between the cadet nurses and the students in the General Curriculum on the Interest (M_f) scale, and also between the cadet nurses and the Music Students on the Hypochondriasis (H_s), Depression (D), Interest (M_f), and Schizophrenia (S_c) scales. Note: In former paper, no statistically significant differences on the separate scales were reported.

4. On the basis of this study . . . The Minnesota Multiphasic Personality Inventory is not a useful instrument for differentiating between those who are more suited for one occupation than another except, perhaps, for women desiring to prepare for the nursing profession. Even here, the differences are of such small magnitude, even though statistically significant, as to have little practical guidance or selection import.

Received January 26, 1951.
Early publication.

Table 2, Revised

Comparison of Women Students in Four College Curricula on the Separate Inventory Scales Showing Difference, Standard Error of Difference, and Critical Ratio of T-scores

MMPI Scales	M_1 (Gen. Curr.) M_2 (Music Curr.)			M_1 (Gen. Curr.) M_3 (Nurse Curr.)			M_1 (Gen. Curr.) M_4 (Lib. Arts Curr.)			M_2 (Music Curr.) M_3 (Nurse Curr.)			M_2 (Music Curr.) M_4 (Lib. Arts Curr.)			M_3 (Nurse Curr.) M_4 (Lib. Arts Curr.)		
	Diff.	σ_{M1-M2}	C.R.	Diff.	σ_{M1-M3}	C.R.	Diff.	σ_{M1-M4}	C.R.	Diff.	σ_{M2-M3}	C.R.	Diff.	σ_{M3-M4}	C.R.	Diff.	σ_{M3-M4}	C.R.
?	.00	.47	.00	.74	.58	1.28	.29	.55	.53	.56	.58	.97	.31	.50	.62	− .25	.64	− .39
L	.00	.53	.00	.09	.60	.15	− .10	.63	− .16	.09	.60	.15	− .10	.62	− .16	− .19	.69	− .28
F	.13	.75	.17	− .18	.89	− .20	− .10	.90	− .11	− .31	.81	− .38	− .18	.83	− .22	.13	.95	.14
H_s	−2.37	1.20	−1.97	1.17	1.21	.97	−1.69	1.52	−1.11	3.54	1.09	3.25	.68	1.42	.48	−2.86	1.43	−2.00
D	−2.22	1.62	−1.37	1.29	1.68	.77	−1.08	1.86	− .58	3.51	1.22	2.88	1.14	1.46	.78	−2.37	1.51	−1.57
H_y	− .38	1.35	− .28	.26	1.38	.19	− .01	1.56	− .01	.64	1.22	.52	.37	1.41	.26	− .27	1.45	− .19
P_d	1.16	1.78	.65	1.60	1.82	.88	− .14	2.07	− .07	.44	1.42	.31	−1.30	1.69	− .77	−1.74	1.74	.00
M_f	.99	1.39	.71	−4.21	1.60	−2.63	−1.92	1.86	−1.03	−5.20	1.47	−3.54	−2.91	1.75	−1.66	2.29	1.92	1.19
P_a	− .88	1.30	− .68	2.63	1.48	1.78	− .08	1.71	− .05	3.45	1.38	2.50	.74	1.62	.46	−2.71	1.77	−1.53
P_t	2.00	1.40	1.43	3.40	1.79	1.90	3.10	1.56	1.99	1.41	1.42	1.00	1.11	1.49	.75	− .30	1.58	− .19
S_c	1.05	1.40	− .75	2.86	1.50	1.91	1.86	1.57	1.18	3.91	1.35	2.90	2.91	1.43	2.03	−1.00	1.53	− .65
M_a	.06	1.44	.04	− .39	1.55	− .25	− .01	1.57	− .01	− .45	1.62	− .28	− .07	1.63	− .04	.38	1.73	.22

Ortho-Rater Color Vision Test Compared with the Ishihara and the Pseudo-Isochromatic Plates

Newell C. Kephart and Melvin J. Tieszen

Division of Applied Psychology, Purdue University

Tests of color vision utilizing pseudo-isochromatic plates provide a simple and rapid means for detecting the presence of defective color vision. Two such tests, the Ishihara and the Pseudo-Isochromatic Plates are used extensively in industry as well as in other fields. An additional test utilizing pseudo-isochromatic plates which is used in industry is the color vision test in the Ortho-Rater.[1]

Purpose of Investigation. While the Ishihara test and the Pseudo-Isochromatic Plates have become accepted as "valid" detectors of defective color vision, the particular function tested by the Ortho-Rater color vision test is in question (**1,6**). The purpose of this investigation is to compare the performance on the color vision test in the Ortho-Rater with performance on the Ishihara test and the Pseudo-Isochromatic Plates.

The seventh edition of the Ishihara test was used throughout the present investigation. This edition consists of 32 plates, of which the 24 intended for the testing of literate subjects were used. Hardy, Rand, and Rittler (**3**) have identified 18 plates from the Pseudo-Isochromatic series which they found to be a good screening device for color defectives.

The color vision test in the Ortho-Rater is the F-7 slide situated at the optical equivalent of eight meters. This slide is a Kodachrome transparency of four Ishihara plates (**6**). Two of the plates contain a single digit, while two are two-digit plates. The score is the number of separate digits correctly responded to, a "perfect" score being six. It is stated in the manual accompanying the Ortho-Rater (**10**, p. 6) that, "The present slide is temporary and classifies subjects finely only at the low levels of color discrimination ability." No claims are made that the test differentiates types of color defectiveness.

[1] The Ortho-Rater is a visual testing device manufactured by the Bausch and Lomb Optical Company for the visual classification and placement of industrial employees.

Standard procedure was used in administering the three sets of tests to two groups, one composed of 110 and the other of 52 male college students. The purpose of giving the tests to a second group was twofold: (1) to obtain an estimate of the reliability of the original comparison; and (2) to cross-validate a proposed scoring key on the Ortho-Rater color vision test.

The Ishihara test and the Pseudo-Isochromatic Plates were used as criteria for the classification of subjects as color normal and color defective. The performance scores of the two groups thus classified on the Ortho-Rater color vision test were compared. A revised scoring of the responses on the Ortho-Rater tests based solely on a statistical analysis of the responses made by the color normals and color defectives will be considered in the results.

Results for Group I

Performance Scores on Tests. Presented in Table 1 [2] is a tabulation of the number of subjects responding correctly to various numbers of the 24 plates in the Ishihara test. It is shown that 14 subjects failed to respond correctly to as many as 14 plates and so were classified as color defective.

Shown in Table 2 are the number of subjects making correct responses to the 18-plate selection of the Pseudo-Isochromatic Plates. Failure to respond correctly to five or more plates on this abbreviated test is considered indicative of defective color vision. Accordingly, 14 subjects are classified as color defectives. These 14 subjects are the same subjects so classified on the Ishihara test.

[2] Tables 1, 2, 3, 4, 5, and 6 and Figures 1, 2, 3, and 4 present the detailed data. These tables have been deposited with the American Documentation Institute. Order Document 2917 from American Documentation Institute, 1719 N St., N.W., Washington 6, D. C., remitting $0.50 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $0.50 for photocopies (6 × 8 in.) readable without optical aid.

The number of subjects earning performance scores ranging from 1 to 6 on the Ortho-Rater color vision test is shown in Table 3. It is apparent from the distribution that no dichotomous grouping has resulted. If the critical score were set at the performance score of two, that is, if those who failed to respond correctly to more than two digits were classified as color defectives, 71 per cent, classified as color defectives by the Ishihara and Pseudo-Isochromatic Plates would be correctly classified; five per cent, classified as color normals by the other two tests, would be misclassified. Setting the critical score at the performance of three, 100 per cent, classified as color defective by the criterion tests, would be correctly classified; ten per cent, classified as color normals, would be misclassified.

Revision of Ortho-Rater Scoring Key. Examination of responses to the digits on the Ortho-Rater test reveals that 100 per cent of the color normal group and 29 per cent of the color defective group read the correct response of five on the first plate. On the second plate, 100 per cent of the color normal group and 93 per cent of the color defective group made the correct response of eight. Eighty-two per cent of the color normal group made the correct response of five to the first digit of the third plate, while only 15 per cent of the color defective group made this response. - Seventy-one per cent of the color defective group could read no digit on the third plate. Twelve per cent of the color normal group could read no digit on the third plate. Twelve per cent of the color normal group read the digit six instead of five, while none of the color defective group made this response. Seventy-seven per cent of the color normal group made the correct response of six to the second digit on the third plate. Fourteen per cent of this group read the digit eight in place of six, while none of the color defective group made this response. On the fourth plate 82 per cent of the color normal group made the response of seven to the second digit on the fourth plate, 27 per cent of this group read two for seven. Seventy-one per cent of the color defectives failed to see any digit on plate four, and none made the response of two for seven for the second digit.

Since some of the "wrong" responses made by those classified as having normal color vision by the Ishihara test and the Pseuo-Isochromatic Plates appear to be typical and are not common to the responses made by the color defective group, it is possible to devise a scoring key that will better differentiate between the color normals and color defectives. It is suggested that the scoring of plates one and two be continued as it is now. In plate three it is suggested that six be an alternate answer for five and that eight be an alternate answer for six. In plate four it seems advisable that the two be scored as is now done, but that two also be an alternate answer for seven.

The acceptance of three additional digits as correct or normal responses in plates three and four would produce the change in results graphically presented in Figures 1 and 2. The number of color normals making a perfect performance score of six increases from 32 to 61 per cent. Establishing a critical score of three would result in 100 per cent of the color defectives being correctly classified. Three per cent of the color normals would be misclassified as color defectives instead of the ten per cent misclassified by the scoring key presently used.

Results for Group II

Performance Scores on Tests. In Table 4 it is shown that 11 of the 52 subjects composing the second group responded correctly to less than 60 per cent of the Ishihara plates and so are classified as color defectives.

The same eleven subjects were placed in the color defective classification by the Pseudo-Isochromatic Plates. As for the original test group, there was complete agreement in the classification of subjects by the Ishihara test and the Pseudo-Isochromatic Plates.

The tabulation of the number of subjects passing from one to six of the digits in the Ortho-Rater color vision test is given in Table 6. As in the previous group, no subject failed all six digits. It is again necessary to determine a critical score that will best separate the color defectives from the color normals. Setting the critical score at a performance score of two, 91 per cent, classified as color defectives on the Ishihara test and the Pseudo-Isochromatic Plates would be correctly classified; 2 per cent, classified as having normal color vision by the two criterion tests, would be misclassified. Setting the critical score at a performance score of

three, 91 per cent, classified as color defectives on the Ishihara test and the Pseudo-Isochromatic Plates, would be correctly classified; 17 per cent, classified as color normals on the criterion tests would be misclassified. Whereas 71 per cent of the color defective subjects in Group I had performance scores of two or less, 91 per cent of the color defectives in Group II fall at these performance scores; the other nine per cent earn a score of four. Ten per cent of the color normals of the first group earned a performance score of three or less, whereas 17 per cent of the color normals in the second group fell below a performance score of three.

Revision of Ortho-Rater Scoring Key. An analysis of responses to the digits on the Ortho-Rater test reveals that a like or greater percentage of color normals in Group II gave the responses of six for five and eight for six in plate three and two for seven on plate four than did the color normals in Group I. None of the color defectives in Group II made these specific responses. The effect of the proposed revision of scoring on the distribution of the second group's performance scores is graphically presented in Figures 3 and 4. If a performance score of three is accepted as the critical score, 17 per cent of the color normals and nine per cent of the color defectives would be misclassified by the present method of scoring. The revised key results in four per cent of the color normals and nine per cent of the color defectives in Group II being misclassified.

Summary

Two groups of male college students, one containing 110 subjects and a second containing 52 subjects were administered the Ishihara test, the Pseudo-Isochromatic Plates, and the color vision test in the Ortho-Rater.

The results of testing the group of 110 subjects showed perfect agreement in the classification of normal and defective color vision on the Ishihara test and the Pseudo-Isochromatic Plates when criterion scores developed by Hardy, Rand, and Rittler (4) were used. Establishment of a critical score at a performance score of three (reading correctly three of the six digits) on the Ortho-Rater color vision test resulted in 100 per cent of the subjects with defective color vision and ten per cent of the subjects with normal color vision falling below this critical score when the classification by the other two tests is accepted. A revised scoring key resulted in a reduction from ten per cent to three per cent in the number of subjects with normal color vision who were misclassified; the scores of subjects with defective color vision remained unchanged.

The Ishihara test and the Pseudo-Isochromatic Plates agreed in the classification of defective color vision and normal color vision on the second group. Ninety-one per cent of the subjects with defective color vision and 17 per cent of the subjects with normal color vision, as classified on the basis of the other two tests, fell below a critical score of three on the Ortho-Rater color vision test. The proposed new Ortho-Rater key reduced the number of misclassified subjects with normal color vision from 17 per cent to four per cent; the scores of the subjects with defective color vision remained unchanged.

Received May 11, 1950.

References

1. Boice, Mary L., Tinker, M. A., and Paterson, D. G. Color vision and age. *Amer. J. Psychol.*, 1948, **61**, 520–526.
2. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans, Green and Co., 1937.
3. Hardy, L. H., Rand, Gertrude, and Rittler, M. Catherine. Tests for the detection and analysis of color-blindness: the Ishihara test. *Opt. Soc. Amer. J.*, 1945, **35**, 268–275.
4. Hardy, L. H., Rand, Gertrude, and Rittler, M. Catherine. A screening test for defective red-green vision. *Opt. Soc. Amer. J.*, **1946, 36,** 610–614.
5. Ishihara, S. *Tests for colour-blindness.* (7th Ed.) Tokyo: Kanehara, 1936.
6. Jobe, F. W. Instrumentation for the Bausch and Lomb industrial vision service. *Bausch and Lomb Mag.*, 1944, **20**, 6–7, 14–15.
7. Tiffin, J., and Kuhn, Hedwig S. Color discrimination in industry. *Arch. Ophthal.*, 1942, **28**, 851–859.
8. *Pseudo-isochromatic plates for testing color perception.* Southbridge, Mass.: Amer. Opt. Co., 1940.
9. *Report on defective colour vision in industry.* Committee of the colour group, The Physical Soc. (London). London: Taylor and Francis, Ltd., 1946.
10. *Standard practice in the administration of the Bausch and Lomb occupational vision tests with the ortho-rater.* Rochester, N. Y.: Bausch and Lomb, 1944.

A Study of Tachistoscopic Vision

Newell C. Kephart and James D. Pecsok

Occupational Research Center, Purdue University

Although many studies have been made upon the span of apprehension as measured tachistoscopically, and the effects of tachistoscopic training upon this apprehension, none could be found referring to the effects of tachistoscopic exposure upon perception. Since many industrial jobs today, particularly those involving inspections, require split-second perception, it seems important that some studies be made in this field. The present paper does not attempt to discover the loss of perceptive power upon tachistoscopic exposure, but is merely the first step in an attempt to determine the correlation between a standard vision test (in this case the Ortho-Rater) and tachistoscopic scores (at $\frac{1}{5}$ second exposure).

The Experiment

Equipment. The equipment used included, in addition to such items as chairs and a table, two white screens to be used for the background; a set of checkerboard targets calibrated at 20/80, 20/67, 20/48, 20/40, 20/34, 20/28, 20/24, 20/20, and 20/17; a half-silvered mirror; an electronic timer; and an Ortho-Rater. This equipment was set up as shown in Figure 1, with the half-silvered mirror making an angle of 45 degrees with the table edge, and the screens adjusted so that the distance from the subject's eye to either screen was 20 feet. Light was furnished by Lumiline bulbs and was equated at 28 foot-candles on each screen. The timer was connected so that light would normally be on only Screen A, the only exception being the interval after the time switch was thrown ($\frac{1}{5}$ second here), at which time it would be on Screen B only. Through the half-silvered mirror, the subject could see only the screen which was lighted. The checkerboard targets were pasted on 8 inch squares of white cardboard, with the target edge parallel to the edge of the cardboard.

Subjects. The subjects were students in the beginning course of Psychology at Purdue University. In most cases they were tested with either eye, so that 108 results were obtained from 46 males and 15 females.

Method. All subjects were first given the complete battery of Ortho-Rater tests in the prescribed form, except that all monocular measurements were made with the other eye occluded. After the completion of the tests they were asked to sit so that their chin was on the chin rest, and the peep hole was adjusted so as to permit them to see the screen with one eye only. They were told to fixate with one eye upon an 8 inch square target on Screen A. This square was identical with those on which the targets had been pasted, and was so placed that a fixation point at its center would appear to be in the same place as the center of the checkerboard targets.

The following instructions were then given: "I am going to flash on the screen targets exactly like the ones you just saw in the Ortho-Rater. The checkerboard will always be on the top, bottom, right, or left, never in the center. Please call out where you see the checkerboard." In order to give them an idea of what was to come, the largest, or 20/80, target was placed on Screen B and shown once. If the subject failed to understand what was expected, it was shown again. In almost every case there was no trouble, however, and the target was shown nine more times, in the order R R T R L B L T B. These showings were made as rapidly as the target could be manipulated, with the subject being given a warning of approximately one second before each showing.

After a rest of thirty seconds, the second largest, or 20/67, target was shown ten times, also as rapidly as possible. This sequence of thirty seconds rest and then ten showings of the target was continued, using the targets in order of largest to smallest, until the subject failed to get at least five correct answers. (Five was arbitrarily decided to be the passing point after talks with several of the subjects who had gotten four correct responses had convinced the experimenter that they were guessing.)

The order of presentation of the targets was as follows:

Target	Order	Target Classification
1	T R R T R L B L T B	20/80
2	L B L T B T R R T R	20/67
3	B T T R L R B T R L	20/48
4	R B T R L B T T R L	20/40
5	T R R T R L B L T B	20/34
6	L B L T B T R R T R	20/28
7	B T T R L R B T R L	20/24
8	R B T R L B T T R L	20/20

The sequence for targets 1 and 3 was picked by cutting a deck of cards, that for targets 2 and 4 is the same except that the starting point is presentation No. 5 instead of No. 1. Targets 5, 6, 7, and 8 were presented in the same order as were 1, 2, 3, and 4.

Scoring. Since, as previously mentioned, it was decided to make five correct responses the minimum for which any credit was to be given, it was necessary to derive a formula that would give four correct responses a score of zero. In order to accomplish this, the empirical formula $X = \frac{(R - 2/3\,W)\,10}{100}$ was developed, where R is the number of correct responses and W is the number of incorrect responses (counting no response at all as being incorrect). The figure thus obtained was applied to the difference between the denominators of the target at which a miss occurred and the next larger target, and the result was subtracted from the larger number to give the score on the smaller target. As an example, suppose the following to be the scores of one subject:

Target	20/80	20/67	20/48	20/40	20/34	20/28
No. Rt.	10	10	9	7	5	4

Through the 20/67 target the subject would get full credit, or a score of 67. For the 20/48 target his score would be $\frac{(9 - (2/3)1)\,10}{100} \times 19 = 16$. This number, when subtracted from 67, would give him a score of 51 through the 20/48 target. For the 20/40 target his score would be $\frac{(7 - (2/3)3)\,10}{100} \times 8 = 4$. This, subtracted from 51, gives him a score of 47 through the 20/40 target. Similarly, for 20/34 his score is $\frac{(5 - (2/3)5)\,10}{100} \times 6 = 1$, giving him a total score of 46. The 20/28 target is disregarded, as the credit is zero.

Since, in the Snellen system, the difference between 20/50 and 20/55 is not the same as the difference between 20/30 and 20/35, it was decided to change these scores to percentage of visual efficiency by using the table on page 126 of Tiffin (8). It must be remembered, however, that these scores are relative, and not absolute. That is, the fact that a subject emerged with a score of 70 does not mean that he is only 70% efficient under the conditions of the experiment.

After the efficiency scores had been obtained in this manner they were grouped by threes, and these grouped scores were used in all calculations.

Results

The correlation between the tachistoscopic scores and the Ortho-Rater scores for vision

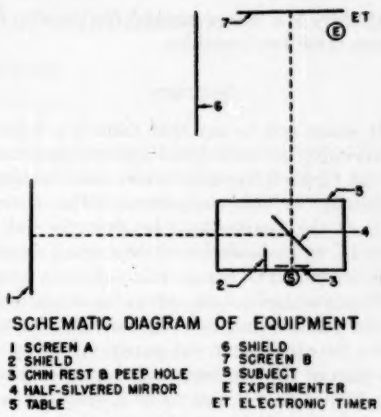

FIG. 1. Diagram of experimental room.

with one eye at distance was calculated to be $.68 \pm .05$. The generally accepted figure for the reliability of the Ortho-Rater is .865. A test-retest reliability check gave .81 as the reliability of the tachistoscope test. Using .81 as the reliability of the tachistoscopic tests, and .865, as the reliability of the Ortho-Rater, the above correlation when corrected for attenuation is .81. The results in Figure 2 show the mean tachistoscopic score for individuals making each Ortho-Rater score. It can be seen

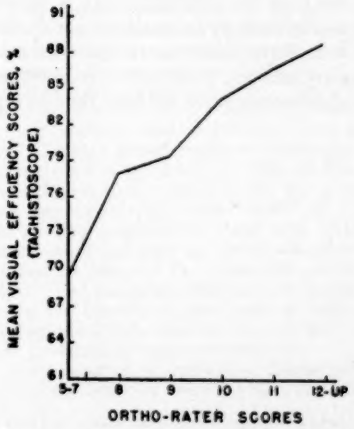

FIG. 2. Tachistoscopic scores *vs.* Ortho-Rater acuity scores (one eye at distance).

that there is a rather marked relationship between these two variables.

Summary

It seems safe to say that there is a definite relationship between visual acuity as measured on the Ortho-Rater and as measured tachistoscopically in this experiment. The critical ratio of the Pearsonian *r* between the two is over 13, which is significant beyond all reasonable doubt. The actual value of the *r* seems still open to doubt, however, as the sample used in the present study was undoubtedly biased. Since the experiment was purely voluntary on the part of the subjects, and since they knew in advance that it was to be a vision experiment, it is probable that there was some preselection in that a number of those with poor vision simply did not enter into it. A still greater cause of doubt was necessitated by the limitations of the target used. The largest target available was calibrated at 20/80 and there is a segment of the population which is unable to see that well even under normal conditions, let alone under those which existed during the test. The results of subjects who were unable to see the first target were omitted from the calculations. Since the Ortho-Rater scores for these subjects were invariably low, such omissions would tend to give a spuriously low correlation figure.

In view of the relatively high *r* between tachistoscopic acuity as measured in this study and Ortho-Rater acuity scores obtained in the customary manner, it does not appear that any marked difference exists between the two types of acuity measurement. It is therefore felt that, for industrial purposes, adequate acuity tests administered in the customary manner will give satisfactory results in selection and placement of personnel and that the addition to a vision test battery of tachistoscopic measures of acuity would be of questionable value.

Received May 17, 1950.

References

1. Adams, J. K. *A test-retest reliability study of the Bausch and Lomb Ortho-Rater with Navy personnel.* OSRD 3969, 1944, Publ. Bd., No. 18324, U. S. Dept. of Commerce: Washington, D. C., 1946.
2. Aveling, F. The perception of tachistoscopically exposed symbols. *Brit. J. Psychol.*, 1932, **22**, 192–199.
3. Ayers, A. W. *A comparison of certain visual factors with the efficiency of textile inspection.* Ann Arbor, Mich.: University Microfilms, 1941.
4. Fry, Glenn A. The significance of visual acuity measurements without glasses. *Amer. J. Optom.*, 1948, **25**, 199–210.
5. Imus, H. A. Comparison of the Ortho-Rater with clinical ophthalmic examinations. *Amer. Psychologist*, 1946, **1**, 283–284.
6. Jobe, F. W. Instrumentation for the Bausch and Lomb industrial vision service. *Bausch and Lomb Mag.*, 1944, **20**, 6–7, 14–15.
7. Shepard, C. F. Visual acuity notations. *Optom. Wkly.*, 1943, **33**, 1365–1366.
8. Tiffin, J. M. *Industrial psychology.* New York: Prentice-Hall, 1947.
9. Zachert, Virginia. A factor analysis of vision tests. Unpublished Ph.D. thesis done at Purdue University, 1949.
10. *Standard practice in the administration of the Bausch and Lomb occupational vision tests with the Ortho-Rater.* Rochester: Bausch and Lomb, 1944.

An Experimental Consumer Panel Technique

Edwin A. Fleishman

The Ohio State University

Experimental investigations of product preferences are usually confined to the laboratory situation with a restricted sample or to simple spot tests made in conjunction with consumer surveys. Previous experiments of these types have been made with shaving creams (**11**), breads (**6**), cigarettes (**1, 4**), and cola drinks (**5, 7, 8, 9, 10**). In an experiment by the writer (**3**), preferences for a product were allowed to develop over a period of time. The study reported here was a further attempt to obtain a larger sample of behavior regarding consumer preferences than usually results from traditional techniques. Moreover, the experimental investigation was conducted at the consumer level, using a panel of families selected from a cross-section of a large metropolitan city. The products investigated included six well-known brands of bottled beer.

The Problem

The purpose of the study was to observe the formation of beer preferences during a seven-day period, under conditions in which all means for identifying the beers except actual brand names were available to the subjects.

The Method

The procedure included: (1) the selection of an unbiased panel of twenty families; (2) supplying them with unlabeled bottles of beer for a seven day period; and (3) observing and recording their preferences during that time.

The Consumer Panel. The panel of beer consumers included families living in different sections of the city. The same proportion of families from each of the four socio-economic groups was included in the sample as is found in the general beer-drinking population of that city.[1] Fifteen per cent of the families were Negro. In the sample, were beer drinkers whose expressed preferences included many different brands. The number of families whose previous preferences included each of the brands in the study was equated.

An important factor in selecting the families (as in all panel techniques) was their availability and their willingness to cooperate. This had been determined during a previous survey by the writer involving 250 respondents. The families within the classifications mentioned above were then selected randomly. The potential subjects were reinterviewed to assure their cooperation and their understanding of the study.

Procedure. The experiment lasted for seven days, during which time the families drank as many bottles of beer as they wanted. The study was conducted during the hot summer months when consumption was at a maximum.

The experimental bottles of beer were provided the families each morning. Each day each home received 48 bottles of beer in blank cases. Included were eight bottles of each of six different brands from which they could choose during the rest of the day. The bottles contained no labels or brand names. Instead, the tops on the bottles of each brand were painted with different colors. Thus, the tops of one brand might all be painted red, the tops of a second brand might all be green, and so on. All beers of the same brand would have tops of the same color on any one day. The subjects were told that the colors would be reassigned to the brands by random selection from day to day. Thus, a brand might or might not have the same color top from one day to the next. The subjects had to make a new set of choices each day as to what they "liked" or "didn't like." The subjects were not told which six brands were included in the experiment. All the bottles were of the same size, shape and color. The subjects could not distinguish between the brands on the basis of beer color. During the study they drank no beer except that provided by the experimenter.

The subjects kept records of the brands they drank (according to the top color) on the check

[1] In a previous study (**2**) it was found that the "A" socio-economic group consumed 4.2% of the beer sold in this city, the "B" group consumed 7.3%, the "C" group 38.1%, and the lowest group, "D," consumed 50.4%. The panel contained approximately these ratios of families within each socio-economic group. Thus, of the 20 families selected, one was from the "A" group, two from the "B" group, seven from the "C" group, and the remaining ten from the "D" group.

form provided each day. They also expressed their preferences for that day on the space provided on the form. Each day the remaining bottles, bottle caps, and empties were collected, the forms checked against the bottles returned, and the home provided with the next day's supply.

The subjects were provided with more than they needed so that they could have an unrestricted choice of brands. On the weekend they were provided with twelve of each brand instead of the customary eight. During the course of the study frequent follow-up calls were made to insure that instructions were being followed and that there were no misunderstandings. The families were also provided with a set of printed instructions.

The study operated on the basic assumption that the people would drink more of what they liked and tend to avoid more of brands they disliked. Since the colors on the six brands were changed from day to day, a new set of decisions was made each day. The data offer two indices of preference. One includes the differences between the brands in the number of bottles of each consumed during the seven day period. The other is represented in the day by day tabulation of the brands (according to top color) named as preferred by the individuals in the study.

Results

The results show that there were significant differences between brands in the number of bottles of each consumed by the panel of consumers. Table 1 presents the total number of bottles of each brand consumed during the seven day period and the percentage of the total this represents for each brand.

Table 1

Absolute and Relative Frequency with Which Bottles of Each Brand of Beer Were Consumed During the Seven Day Period

Brands	Number of Bottles Consumed	Percentage of Total
A	625	18.4
B	613	17.9
C	591	17.4
D	566	16.6
E	514	15.1
F	497	14.6
Total	3406	

Table 2

Percentage of Times Each Brand Was Named as "Liked Best" During the Course of the Study

Brands	Percentage "Liked Best"
A	23.2
C	21.4
B	18.7
D	17.1
E	11.9
F	7.7

The chi square based on Table 1 was 16.7 which is significant beyond the .01 level. In other words, there appeared to be real differences in preference for beers even when the brand names were not known. A more complex analysis, which involved successive elimination of different brands and combinations of brands from the analysis revealed that the majority of the variation was a result of the avoidance of Brands E and F. The differences between the number of bottles of Brands A, B, and C consumed were not significant. Brand D was not preferred significantly more than the last two brands and was not avoided significantly less than the leading three brands in the study.

In addition to drinking most of what they liked, all the subjects were requested to write in on the daily record sheet they received, the color of the top on the brand they liked best that day. The number preferring each "color" was tabulated under the appropriate brand each day. Table 2 summarizes the proportion of times each brand was named as preferred over the remaining five during the seven days. In all, 327 choices were made.

The relationship between expressed preference and the amount of each brand actually consumed is shown to be high.

An analysis of the day to day variation in the number of bottles of each brand consumed daily yielded a chi square which was not significant (chi square = 22.4, d.f. = 25, $P > .05$). Thus, although there were significant differences between brands in the total number of bottles of each consumed, there was little shift from day to day in the brands preferred and avoided. The trend was generally maintained through

the week. The two avoided brands were consistently avoided, and the brand high each day generally varied among the three preferred brands.

An analysis of differences in preference between the families in the panel, however, showed wide inter-family differences in the number of bottles of each brand consumed. Thus, every brand was preferred by some family during the study. However, more of the families formed preferences for Brands A, B, and C, than for the other brands.

It was also found that preferences formed were for the most part unstable and most subjects did not actually pick out the same brands from day to day. In other words, the individuals contributing to the total number of bottles consumed of a particular brand during the week, differed from day to day. Thus, although more of the subjects liked Brands A, B, or C, more of the time and avoided Brands E and F more of the time, individual preferences generally tended to shift from day to day.

The relationship between previous brand preferences and preferences formed during the experiment was also investigated. The percentage of times Brand A drinkers said they liked Brand A during the study, the percentage of Brand B drinkers who named Brand B (by colored top, of course), and so on, was as follows: Brand A drinkers, 56.3 per cent; Brand D drinkers, 45.2 per cent; Brand C drinkers, 28.9 per cent; Brand B drinkers, 26.3 per cent; and Brand F drinkers, 12.5 per cent. Thus, users of Brands A and D seemed more apt to form preferences for their old brands when the brand names were not known than did users of other brands. The percentage of times drinkers of Brands B, C, and F picked their own brands are within the limits of chance expectancy.

Summary

A panel of consumers was selected from a large city and their preferences for brands of beer investigated through the use of an experimental technique.

1. The study showed that there were real differences in preferences for beers even when the brand names were not known.

2. Of the beers investigated, three were generally preferred, one occupied a position just below the leaders, and two were more generally avoided. This was on the basis of the total number of bottles consumed during the study.

3. The preferences expressed by the subjects during the study were in agreement with the amount actually consumed.

4. The general trend in preference for the families in the study generally prevailed for the group from day to day.

5. Preferences expressed by the individuals in the study generally proved to be unstable and shifted from one brand to the next. However, more people preferred three of the brands and avoided two others more of the time.

6. However, previous users of two of the brands seemed better able to form preferences for their old brands when the brand names were not known than users of the other brands.

7. The study indicates that important information may be gained using experimental techniques at the consumer level. The fruitfulness of this approach to problems of market research should increase with more refined techniques and research projects of longer duration.

Received May 6, 1950.

References

1. Chase, S. Blindfolded you know the difference. *New Republic*, 1928, **55**, 296–298.
2. Fleishman, E. A. An evaluation of consumers' beer preferences. Unpublished study, 1949.
3. Fleishman, E. A. An experimental study of cigarette preference. Paper read at the Eastern Psychological Association, Springfield, Mass., April, 1949.
4. Husband, R. W., and Godfrey, J. An experimental study of cigarette brand identification. *J. appl. Psychol.*, 1934, **18**, 220–223.
5. Jenkins, J. G. An application of psychological techniques to market research. *Psychol. Bull.*, 1936, **33**, 726.
6. Katz, D. A study of the taste of bread. *Hum. Factor, Lond.*, 1937, **16**, 241–246.
7. Pronko, N. H., and Bowles, J. W., Jr. Identification of cola beverages: First study. *J. appl. Psychol.*, 1948, **32**, 304–312.
8. Pronko, N. H., and Bowles, J. W., Jr. Identification of cola beverages: A further study. *J. appl. Psychol.*, 1948, **32**, 559–564.
9. Pronko, N. H., and Bowles, J. W., Jr. Identification of cola beverages: A final study. *J. appl. Psychol.*, 1949, **33**, 605–608.
10. Pronko, N. H., and Herman, D. T. Identification of cola beverages: Postscript. *J. appl. Psychol.*, 1950, **34**, 68–69.
11. Schlosberg, H. A well-controlled method of evaluating consumers' goods. *J. appl. Psychol.*, 1941, **25**, 401–407.

Book Reviews

Hattwick, M. S. *How to use psychology for better advertising.* New York: Prentice-Hall, Inc., 1950. Pp. xvi + 376. $5.75.

In a stimulating and entertaining book Dr. Hattwick shows that the application of the findings of psychology can lead to the successful practice of advertising. Although the book is written in popular, almost telegraphic style, the author includes countless experimental studies and carefully chosen illustrations to indicate the fruitful utilization of scientific methods in a field which is still fluid in terms of methodology as well as findings.

The book is divided into four major parts and twenty-one chapters. Part 1 (2 chapters) is entitled "What Do Consumers Want?" and concerns itself with a discussion of human needs and means of their satisfaction. Part 2 (14 chapters) is called "Six Steps That Make Advertising More Effective," and deals with problems of consumer research, the make-up of the market, finding the right appeal, attention-getting devices, keeping the consumer's interest and impressing his memory. Part 3 (3 chapters) discusses "Methods of Evaluating Advertising Effectiveness," and Part 4 (2 chapters) concludes the book with a discussion on "Building Customer Confidence in Advertising." Concise summaries and selected references are included at the end of each chapter.

The pleasant, rambling organization, perhaps one of the greatest assets of the book, also offers drawbacks, especially from the view-point of the teaching psychologist. There might be more theoretical discussions of motivation, emotion, attention, perception or learning, although pertinent material from all of these areas can be found throughout the book. There is no adequate treatment of the assumptions of the scientific method, although this type of discussion would certainly have a place in a book designed to sell science to the practitioner. There are some inaccuracies in psychological theory, such as a listing of a specific number of basic human wants, or a statement that "practically all persons . . . have the same fundamental drives (wants) that are usually satisfied in about the same ways" (p. 24).

All in all Dr. Hattwick's book should be highly appreciated by the practical advertiser who is not familiar with the contributions of psychology to the knowledge in his field; the teacher of psychology will find it very useful as an up-to-date basic text, provided he presents some of the theoretical material from other sources.

Irving R. Weschler

University of California, Los Angeles

Harrel, Thomas Willard. *Industrial psychology.* New York: Rinehart and Company, Inc., 1949. Pp. xvii + 462. $4.25.

The objectives of this book are clearly indicated in the author's preface when he states that today's industrial psychology has as its primary function the analysis of human relations in industry rather than the use of tests. One has a feeling while reading the first third of the text, however, that the author has not been entirely willing to give himself over to this "new" approach, for he discusses at some length the kinds of tests and their uses. Unfortunately the "uses" sections are often too brief to be really instructive. In several instances the author seems hesitant about recommending validated methods and says "this could be used"; "this might be used." The reviewer feels that users of industrial psychology's methods are ready for more positive suggestions.

One section of the text is titled "Human Engineering." Perhaps this would be a proper time to voice an objection to the use of the term "human engineering" when one speaks of materials, equipment, and environmental factors. Engineering procedures are reflected in human behavior and are appraised in terms of human response. The use of the term by commercial vocational testing groups also adds to the confusion. This leads the reviewer to suggest that industrial psychologists might profit by deciding to use a less confusing term.

Most of this section is concerned with brief descriptions of those areas which constitute

special subject matter fields in advanced psychology and industrial management. The absence of material on equipment design is particularly apparent.

When the author returns to his basic objective of dealing with the analysis of human relations in industry the text makes a definite gain in interest value and practical application of the psychologist's knowledge. The excellent bibliography on attitudes and job satisfaction cannot, however, overcome the over-detailed enumeration of studies and the small emphasis on evaluation of the chapter on that subject. There are many reviews of the kinds of surveys and where they were made but little or no reference is made to the problems solved or the issues met by these surveys. The data of the survey do not comprise the end to be sought, but are only a means for reaching a solution to an industrial problem.

The illustrations used in the discussion of morale and monotony are sometimes hard to follow, which suggests hurried writing and not too careful editing. This comment also applies to some of the tables which are full page and are not clearly delineated as tables. They appear almost as text material.

In the chapter on psychological aspects of labor relations the author fulfills his objective in a most adequate fashion and gives other textbook writers in industrial psychology cause for revising their presentations to make the material more applicable to the real analysis of human relations in industry. It turns attention from the tools of the psychologist to the service he has to offer in the field of industrial problems. This change in perspective is in itself a strong recommendation for the book. No teacher or consultant in industrial psychology can afford to overlook this important point of view.

Eleroy L. Stromberg

Western Reserve University

Deming, W. E. *Some theory of sampling.* New York: Wiley, 1950. Pp. xvii + 602. $9.00.

It is not a criticism of this work to say that only a few psychologists will find it useful, since the volume is mainly concerned with the theory of large-scale sampling of the survey type. The chief emphasis is on distribution theory and the determination of appropriate sampling error formulas that go with various sampling schemes. Those whose calculus is rusty will find the going tough at times, especially in the chapter devoted to the gamma and the beta functions.

The book is not entirely mathematical in character; dispersed throughout one finds many practical suggestions and many examples illustrating the use of various sampling methods. The excellent nontechnical discussion in the first two chapters will repay careful reading by anyone who is interested in the difficulties that beset sampling surveys.

Quinn McNemar

Stanford University

Noland, E. William, and Bakke, E. Wight. *Workers wanted.* New York: Harper and Brothers, 1949. Pp. xi + 233. $3.00.

This study of employers' hiring policies, preferences, and practices in New Haven, Connecticut and Charlotte, North Carolina is the third Interim Report issued under the auspices of the Yale Labor and Management Center.

The study is based on guided interviews held in 1945 and 1946 with the persons responsible for setting hiring policy and directing hiring practice in 135 companies in New Haven and 105 companies in Charlotte. These companies employed approximately eighty per cent of the industrial labor force and represented all major industries (businesses excluded) in the two cities. Interview data were reduced to 99 tables, and the book is essentially a discussion of these tables.

The first two chapters give an overview and a discussion of the order of importance (based on requirements and preferences) assigned by employers to seventeen employee characteristics such as character, personality traits, experience, family status, religion, etc. Each of the next five chapters deals with a particular occupational group: production workers, common laborers, service and maintenance workers, routine clerical workers, and executive and administrative assistants. The next two chapters deal with methods used by employers to determine the extent to which applicants meet requirements and preferences, and sources of potential employees. The last chapter develops a picture of motivations, compulsions, and

organizational requirements of management, and suggests implications regarding the employer, community, and worker. Appendices give a note on methodology and 58 of the tables which form the background of the book.

The data gathering, analysis, and report indicate the close attention to minute detail which is characteristic of sound research. The authors appear to have extracted the last gram of value from their data, whereas many others would have been content with a more superficial and less productive analysis. The book abounds with penetrating insights, interpretations, and speculations.

The book is a popularized interim report, makes reference to only three other publications, contains no bibliography, and contains no index. The result is a highly readable report of value to personnel psychologists, sociologists, economists, industrial executives, union leaders, and others interested in employment.

C. E. Jurgensen
Minneapolis Gas Company

Jennings, Helen Hall. *Sociometry in group relations*. Washington, D. C.: American Council on Education, 1948. 2nd printing 1949. Pp. ix + 85. $1.25.

The quality and effectiveness of human association, both for personal and social ends, has become the subject of scientific scrutiny in many settings. To what extent can the microcosm of the classroom yield answers and originate useful methods for the improvement of social relations?

Jennings and her associates in the project in Intergroup Education in Cooperating Schools has supplied a guide for teachers in the use of sociometric methods. Techniques for selecting situations, framing questions, tabulating choices, and constructing sociograms are described in detail. However, the pamphlet is more than a manual, for leads for the analysis of the process back of the group structure are supplied with much insight based on previous experiment and experience. Suggestions for interviews, for individual treatment, for structural changes, and for the appraisal of growth through successive sociograms are given. Illustrations are presented of problems arising from social, race, and religious differences and of promising ways for improvement.

While the report is based on classroom settings the range of implications and applications is over the entire field of interpersonal relationships.

Willard C. Olson
University of Michigan

Hamrin, Shirley A., and Paulson, Blanche B. *Counseling adolescents*. Chicago: Science Research Associates, Inc., 1950. Pp. 380. $3.50.

This volume, the first book to be published in SRA's new Professional Guidance Service, will probably find itself most appreciated in guidance and education classes in college. It presents the counseling process in a simple, unsophisticated manner without any of the esoteric terminology or circumlocution which so often abounds in this type of literature. In presenting techniques utilized in counseling, the authors claim to be eclectic and state their views in non-technical and easily comprehensible language. They have given us an elementary introduction to the field and have personalized a good portion of the content, the words of wisdom being put into the mouths of two paragons of psychological virtue. The authors employ a common sense approach to the needs of adolescents and to their various pitfalls on the road to maturity and in so doing, they introduce wholesome ideas and attitudes in their own natural and unaffected way. Their treatment of psychological needs and mechanisms is an unpretentious one with emphasis being laid on bringing about an integrated adjustment in both the individual and environment rather than on altering one to suit the other. Each point is well demonstrated by the two fictionalized counselors who reappear through most of the chapters.

The authors consider counseling as a specialized function of school preparation and feel that counseling techniques should be specifically designed for the different age groups. Counselors, it is felt, should understand general psychological concepts, motives, mechanisms, modes of adjustment and specific problems of the adolescent group. Furthermore, the counselor should understand and be able to handle skillfully the counseling process in its various phases. The content of the book deals consecutively with the concept of behavior, the mat-

uration process, philosophies of and attitudes toward counseling, types of problems in the adolescent group and necessary traits for the competent counselor.

The greatest value of this book may be found in its departure from a rigid or stereotyped counseling philosophy and in its freer and more individualized method of dealing with adolescent problems. It should be valuable for undergraduate college classes because of its sound and understandable advice; but its content appears to be too elementary for it to be of any real help to the experienced counselor or teacher. It covers what a counselor needs to know and how he should use this knowledge, but it is not provocative, raises no new issues, and makes no distinctive contribution to the guidance field.

Jewel B. Kaufman

University of Michigan

Britt, Steuart Henderson. *Social psychology of modern life.* (Rev. Ed.) New York: Rinehart & Co., Inc., 1949. Pp. xvi + 703. $4.50.

This revision parallels the organization and content of the first edition and continues to be quite readable and interesting. Its many illustrations, suggested readings, discussion questions commend it for teaching purposes. In certain areas, recent literature has been thoroughly culled, increasing the length by over a hundred pages.

The very extensiveness of the book is one of its shortcomings; touching on so many topics, Britt has not gone deeply into any. The criticism of the first edition—superficiality—still applies. And for this, he has supplied his own best criticism: "After all, he (the student) says, if what we are learning now (facts which may change) may not be true, why should we bother to learn it?" (p. 419). There seem to be too many "facts" presented without enough attention to the methods for obtaining them.

Without claiming to be a systematist, Britt does use certain organizing concepts, such as learning principles and Thomas' four wishes. But, instead of lending meaning to the data, they give a hobby-horse effect: movement without progress.

It is also difficult to tell, despite the first chapter, whether this is a text in general psychology, social psychology, or sociology. The inclusion of so much material on learning, the nervous system, etc., seems unnecessary. The groups with which he deals, such as the church, the gang, and the nation, are sociological classifications. Britt has not made the transfer to the psychologist's interest in the effects on or the origins of these in the individual.

The few actual errors of commission are not serious but should not be in a revision. For example, he cites *without comment* the conclusion of a study in which, because of a very high correlation, it is said that causality may be considered; he calls a series of symmetrical curves normal, though they differ in kurtosis; his principal criticism of the Thurstone technique deals with its difficulty, which is over-exaggerated. While not an error, to refer to E. L. Thorndike as "a world famous psychologist" without naming him seems strange in a text book.

The most serious difficulty, to this reviewer, lies in what has been omitted. Thomsen, reviewing the first edition, found it amazing that such industrial work as that of Roethlisberger and Dickson should have been left out; in the present edition, two short paragraphs seem to suffice. Industrial work is sorely lacking, an unusual situation considering Britt's specialty. To a considerable extent, this is still a study of the social behavior of college sophomores. The uses of propaganda in the hot and cold wars have not affected the treatment of that topic; the major contribution of psychology during the war seems to have been the "Psychologists' Manifesto"; no real changes have been made in the treatment of the totalitarian state; no significant material on men at war has been added.

The student using this book will find it interesting; he will learn many facts about certain aspects of current social life. How much he will learn about the bases for social living, the *psychological* origins of groups, why men behave as they do in their working worlds remains a serious question.

John W. Gustad

Vanderbilt University

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Time and free will. Henri Bergson. New York: The Macmillan Co., 1950. Pp. 252. $2.75.

Modern industrial leadership. Jerald K. Foster Brannon. Deep River, Conn.: National Foremen's Institute, Inc., 1950. $7.50.

The "why" of man's experience. Hadley Cantril. New York: The Macmillan Co., 1950. Pp. 198. $2.75.

Labour relations in London transport. H. A. Clegg. New York: Augustus M. Kelley, Inc., 1950. Pp. 188. $2.50.

Personality and psychotherapy. John Dollard and Neal E. Miller. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 488. $5.00.

Psychological diagnosis and counseling of the adult blind. Wilma Donahue and Donald Dabelstein, Editors. New York: American Foundation for the Blind, Inc., 1950. Pp. 173. $2.00.

The AMA handbook of wage and salary administration. M. Joseph Dooher and Vivienne Marquis, Editors. New York: American Management Association, 1950. Pp. 412. $7.50.

The psychoanalytic study of the child. Ruth S. Eissler, et al., Editors. New York: International Universities Press, 1950. Pp. 410. $7.50.

My job contest. Chester E. Evans and Laverne N. Laseau. Personnel Psychology Monograph No. 1. Washington, D. C.: Personnel Psychology, 1727 Harvard Street, N. W., 1950. Paper, $2.50; cloth, $3.50.

The meaning and practice of psychotherapy. V. E. Fisher. New York: The Macmillan Co., 1950. Pp. 513. $5.00.

The psychology of adolescence. Fourth edition. Karl C. Garrison. New York: Prentice-Hall, Inc., 1950. Pp. 448. $4.50.

Unraveling juvenile delinquency. Sheldon and Eleanor Glueck. New York: The Commonwealth Fund, 1950. Pp. 416. $5.00.

Studies in leadership. Leadership and democratic action. Alvin W. Gouldner, Editor. New York: Harper and Brothers, 1950. Pp. 736. $5.00.

Anxiety. Paul H. Hoch and Joseph Zubin, Editors. New York: Grune and Stratton, 1950. Pp. 254. $4.50.

Studies in prejudice. Five volumes. Max Horkheimer and Samuel H. Flowerman, et al. New York: Harper and Brothers, 1950. $20.00 for all five volumes.

Neurosis and human growth. Karen Horney. New York: W. W. Norton and Co., Inc., 1950. Pp. 391. $3.75.

Repressive forgetting. Ija N. Korner. New York: Bureau of Publications, Teachers College, Columbia University, 1950. Pp. 60. $2.10.

Advertising psychology and research. Darrell B. Lucas and Steuart H. Britt. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 741. $6.50.

The work and training of the psychological therapist. Rollo May. Washington, D. C.: The Psychological Service Center Press, 1950. Pp. 23. $1.00.

An outline of scientific criminology. Nigel Morland. New York: Philosophical Library, Inc., 1950. Pp. 287.

Learning theory and personality dynamics. O. H. Mowrer. New York: The Ronald Press Co., 1950. Pp. 780. $7.50.

Children's reactions to radio adaptations of juvenile books. Mae O'Brien. New York: Columbia University Press, 1950. Pp. 146. $2.00.

Vocational rehabilitation of psychiatric patients. Thomas A. C. Rennie, Temple Burling, and Luther E. Woodward. New York: The Commonwealth Fund, 1950. Pp. 152. $.75.

Principles and procedures in student counseling. Francis P. Robinson. New York: Harper and Brothers, 1950. Pp. 321. $3.00.

Client-centered therapy. Carl R. Rogers. Boston: Houghton Mifflin Co., 1950. Pp. 560. $4.00.

The image and appearance of the human body. Paul Schilder. New York: International Universities Press, 1950. Pp. 353. $4.50.

Twentieth century mental hygiene. Maurice J. Shore, et al. New York: Social Sciences Publishers, 1950. Pp. 444. $6.00.

Residence hall counseling. Calvin S. Sifferd. Bloomington, Ill.: McKnight and McKnight Publishing Co., 1950. Pp. 238. $3.00.

A study of the development of the small college non-academic personnel program. E. Clarendon Smith. Illinois: The College and University Personnel Association, 1950. Pp. 141.

The development of reasoning in children with normal and defective hearing. Mildred C. Templin. Minneapolis: University of Minnesota Press, 1950. Pp. 143. $3.00.

Culture in crisis: a study of the Hopi Indians. Laura Thompson. New York: Harper and Brothers, 1950. Pp. 221. $4.00.

Response of physical systems. John D. Trimmer. New York: John Wiley and Sons, Inc., 1950. Pp. 208. $5.00.

College women who express futility. Pauline Park Wilson. New York: Bureau of Publications, Teachers College, Columbia University, 1950. Pp. 166. $2.75.

The aged and society. A symposium on the problems of an aging population. Champaign, Ill.: Industrial Relations Research Association, 1950. Pp. 237. $3.00.

Training in clinical psychology. Staff of Conference on Graduate Education in Clinical Psychology. New York: Prentice-Hall, Inc., 1950. Pp. 250. $3.00.

Controlling indirect labor and maintenance costs. New York: American Management Association, 1950. Pp. 22. $.50.

Maintaining two-way communication: company experiences and techniques. New York: American Management Association, 1950. Pp. 42. $1.00.

Performance rating and job analysis for sales personnel. New York: American Management Association, 1950. Pp. 27. $.75.

The human relations job of personnel management. New York: American Management Association, 1950. Pp. 40. $.75.

Anderson and Valentine

Reprinted

The article by Anderson and Valentine on "The preparation of articles for publication in the journals of the American Psychological Association," which was originally published in the June 1944 *Psychological Bulletin*, has been reprinted. It is priced at twenty cents a copy, with no discount for quantity orders.

Order from

American Psychological Association

1515 Massachusetts Avenue N.W.

Washington 5, D. C.

March 1951 Best Sellers

of the

American Psychological Association

WOLFLE, HELEN M. Graduate training facilities in psychology: 1951-52. Reprinted from the *American Psychologist*, January, 1951.
28 pages, $.10

DUNCKER, K. On problem solving. Psychological Monograph #270.
113 pages, $2.50

GOLDSTEIN, K., and SCHEERER, M. Abstract and concrete behavior. Psychological Monograph #239.
151 pages, $2.25

PSYCHOLOGICAL MONOGRAPH

#305
1949

THE DEVELOPMENT AND VALIDATION OF A SET OF MUSICAL ABILITY TESTS

BY

ROBERT W. LUNDIN

$1.00

Dr. Lundin's monograph, #305 of the Psychological Monograph Series, is one of the most popular of those recently published.

American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.